

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ENGENHARIA GEOGRÁFICA, GEOFÍSICA E ENERGIA



Extração de Fachadas de Edifícios em Imagens Obtidas com UAV

Carolina Hormigo Mateus Rodrigues

Mestrado em Engenharia Geográfica

Dissertação orientada por:
Professora Doutora Paula Maria Ferreira de Sousa Cruz Redweik

2016

Agradecimentos

Em primeiro lugar, à professora Paula Redweik pelo apoio, disponibilidade e partilha de conhecimentos durante a realização deste trabalho.

Aos meus amigos e colegas que me acompanharam ao longo destes anos, e acompanharão por muitos mais, por tornarem os meus dias melhores e aos meus pais pela oportunidade que me deram.

Por último, agradeço à SINFIC que disponibilizou os dados utilizados neste trabalho.

RESUMO

No âmbito da dissertação final do Mestrado em Engenharia Geográfica, da Faculdade de Ciências da Universidade de Lisboa, foi desenvolvido um algoritmo de reconhecimento de fachadas de edifícios em imagens aéreas verticais, obtidas com UAV (*unmanned aerial vehicle*).

O método proposto tem como dados *input* um conjunto de imagens obtidas com UAV, as respetivas orientações externas (OE) e orientação interna (OI) da câmara e uma ortofoto e um DEM (*Digital Elevation Model*) obtidos preliminarmente no *software* de processamento fotogramétrico de imagens, *PhotoScan*. O algoritmo considera duas etapas principais: Detecção de edifícios e Reconhecimento de fachadas.

Na etapa da detecção de edifícios, o algoritmo recorre aos serviços da *Google*, através das APIs (*Application Programming Interface*) *Reverse Geocoding* e *Geocoding* para obter a localização de um ponto em cada edifício presente na área coberta pelas imagens. São calculadas as coordenadas imagem desse ponto através das equações de colinearidade e, com base no critério ‘maior distância ao centro da imagem’, é identificada a imagem ideal para a extração da fachada.

Na etapa do reconhecimento de fachadas são aplicados alguns operadores de Processamento Digital de Imagem (PDI), como o gradiente, tanto ao DEM, como à imagem da fachada. No primeiro para detetar as posições onde existe probabilidade de existir uma fachada e na segunda para detetar os limites da fachada.

O método foi testado com dados de Sousel, Portalegre, Portugal, no sistema de coordenadas PT-TM06/ETRS89, tendo-se obtido, para alguns casos, percentagens de detecção de fachada e qualidade do método superiores a 90%, mas para outros casos, percentagens inferiores a 50%, o que revela alguma instabilidade no método.

Palavras-chave: UAV, fachadas, Fotogrametria, textura

Abstract

Within the scope of the curriculum project of Geographic Engineering Masters, from Faculdade de Ciências da Universidade de Lisboa, a recognition algorithm of buildings facades in vertical aerial images was developed, which were obtained by UAV (unmanned aerial vehicle).

The proposed method has as input data a set of images obtained by UAV, the photo's exterior orientations (EO) and camera interior orientations (IO), an orthophoto and a DEM (Digital Elevation Model) preliminarily obtained in the photogrammetric image processing software, PhotoScan. The above mentioned algorithm consists in two main stages: Building detection and Facade recognition.

Within the stage of Building detection, the algorithm uses *Google* services through the Reverse Geocoding and Geocoding Application Programming Interface (APIs) to obtain the location of a point in each building present in the area covered by the images. The image coordinates of this point are calculated through the collinearity equations and, based on the criteria 'greater distance to the center of the image', the ideal image for the extraction of the facade is identified.

Within the stage of facade recognition some Digital Image Processing (PDI) operators, such as the gradient, are applied to both the DEM and the facade image. In the first one, to detect the positions where there is a probability of the existence of a facade and in the second one to detect the limits of such facade. In some cases, the facade detection percentage and the quality percentage were higher than 90%, but for other cases it was less than 50%, which reveals some instability in the method.

The method was tested with data from Sousel, Portalegre, Portugal, in the coordinate system PT-TM06 / ETRS89.

Keywords: UAV, facade, Photogrammetry, texture

Índice

1	Introdução	1
1.1	Enquadramento.....	1
1.2	Estado da Arte	1
1.3	Objetivos e contribuição do trabalho.....	2
1.4	Organização do Trabalho	3
2	Fundamentos teóricos.....	4
2.1	UAV	4
2.1.1	Uso de UAV na Fotogrametria.....	4
2.1.2	Viabilidade legal do uso de UAV.....	5
2.2	Geometria da fotografia e projeção central	6
2.2.1	Imagem digital.....	6
2.2.2	Parâmetros que caracterizam a geometria da imagem	6
2.2.3	Correção das coordenadas imagem	7
2.2.4	Efeitos do relevo numa imagem.....	8
2.2.5	Equações de colinearidade	9
2.2.6	Transformação dos 8 parâmetros	10
2.3	Aquisição e processamento de dados com UAV.....	12
2.3.1	Aquisição de imagens e medição de PFs.....	13
2.3.2	Técnica SfM	14
2.3.3	Georreferenciação	16
2.3.4	Nuvem de pontos densa.....	16
2.3.5	Modelo 3D e ortofoto.....	17
3	Dados.....	18
3.1	Descrição da área de estudo	18
3.2	Aquisição de imagens.....	18
3.3	Pontos Fotogramétricos.....	19
4	Operações preliminares	21
4.1	Criação DEM e ortofoto.....	21
4.2	Análise da qualidade dos produtos gerados	23
5	Método proposto	26
5.1	Localização de edifícios	26
5.2	Reconhecimento de fachadas	28
6	Análise da qualidade do método de reconhecimento de fachadas	34

7	Conclusões	39
8	Anexos.....	42
	Anexo I – Exemplos de Algumas fachadas obtidas com o método	42

Índice de Figuras

Figura 1- Comparação de alguns métodos de medição no que diz respeito à exatidão e tamanho do objeto/área. (Fonte: Serra, 2013).....	5
Figura 2- Perspetiva central de uma imagem obtida na vertical e alguns dos parâmetros de orientação interna (distância focal, c , e coordenadas do ponto principal (x_0 , y_0)). (Adaptado de Verhoeven et al., 2013)	7
Figura 3- Origens dos diferentes sistemas de coordenadas foto (Fonte: Redweik, 2011)	8
Figura 4- Distorção radial causada pela topografia e altura dos objetos no terreno. (Adaptado de Verhoeven et al., 2013) por letras terreno grandes	9
Figura 5- Relação entre imagem e objeto (Redweik, 2010)	10
Figura 6- Projeção central de um objeto 3D num plano (Fonte: Redweik, 2008).....	11
Figura 7- Projeção perspectiva de uma reta (Fonte: Redweik, 2008).....	11
Figura 8- Perspetiva entre planos (Fonte: Redweik, 2008)	12
Figura 9- Passos do SfM+MVS. (Adaptado de Verhoeven et al., 2013)	13
Figura 10- Imagens de um cubo com perspetivas diferentes (a) e reconstrução do cubo (b) (Fonte: Agarwal, 2010).....	15
Figura 11- Enquadramento do município de Sousel no contexto nacional e área de estudo, no retângulo	18
Figura 12- Aparelho UAV com o qual foram adquiridos os dados (Fonte: Serra, 2014)	19
Figura 13- Localização das câmaras e sobreposição das imagens dada pelo número de fotografias em que um ponto aparece	19
Figura 14- Distribuição de PFs no bloco	20
Figura 15- Workflow do software Agisoft PhotoScan	21
Figura 16- PF identificado em duas imagens diferentes	22
Figura 17- Parte Sudeste da nuvem de pontos densa com qualidade 'ultra high'	22
Figura 18- Parte Sudeste do DSM obtido através do método TIN (à esquerda) e através do método grid (à direita)	22
Figura 19- Ortofoto	23
Figura 20- Defeitos na ortofoto.....	24
Figura 21- extratos da nuvem de pontos densa.....	24
Figura 22- Elipses de erro para as localizações das câmaras obtidas com GPS/IMU.....	25
Figura 23-Fluxograma da etapa da Localização de edifícios	26
Figura 24- Resultado da aplicação da API para algumas ruas (asteriscos a vermelho).....	27
Figura 25- Fluxograma da etapa do Reconhecimento de Fachadas	28
Figura 26- Gradiente do DEM.....	29
Figura 27- Posições onde existe probabilidade de existir fachadas.....	29
Figura 28- Extrato da ortofoto com posições onde existe probabilidade de existir fachadas (linhas coloridas)	30
Figura 29- Exemplos de edifícios com respetivas posições onde existe probabilidade de existir fachada (linhas coloridas).....	30
Figura 30- Posições dos vértices de uma fachada obtidos pelo cálculo das equações de colinearidade.....	31
Figura 31- Gradiente da intensidade dos pixels para uma fachada	32
Figura 32- Fachada retificada após aplicação das máscaras	32

Figura 33- Exemplo de uma fachada extraída pelo algoritmo	33
Figura 34- Localização das fachadas analisadas.....	34
Figura 35- histograma da distância, em pixels, entre os valores calculados e os valores medidos interactivamente, para as 7 fachadas	34
Figura 36- Imagem de referência (à esquerda) e imagem obtida (à direita)	35
Figura 37- Exemplo de resultado para o qual o método não teve sucesso	36
Figura 38 - Exemplo 2 de resultado para o qual o método não teve sucesso	36
Figura 39- Exemplo de fachada obtida com sucesso	37
Figura 40- Exemplo de fachada obtida com sucesso	37
Figura 41- Exemplo da aplicação de fachadas a um modelo 3D.....	38
Figura 42- Estimativa da posição dos vértices da fachada 1 na imagem (à esquerda) e gradiente da intensidade dos pixels (à esquerda)	42
Figura 43- Exemplo 1 de fachada obtida pelo método	42
Figura 44- Estimativa da posição dos vértices da fachada 2 na imagem (à esquerda) e gradiente da intensidade dos pixels (à esquerda).....	42
Figura 45- Exemplo 2 de fachada obtida pelo método	42
Figura 46- Estimativa da posição dos vértices da fachada 3 na imagem (à esquerda) e gradiente da intensidade dos pixels (à esquerda).....	43
Figura 47- Exemplo 3 de fachada obtida pelo método	43
Figura 48- Estimativa da posição dos vértices da fachada 4 na imagem (à esquerda) e gradiente da intensidade dos pixels (à esquerda).....	43
Figura 49- Exemplo 3 de fachada obtida pelo método	43

Índice de Tabelas

Tabela 1- Contagens de TP, TN, FP e FN e valores obtidos para a pdf, pq e pfn	36
---	----

Índice de Equações

(1) Coordenadas imagem corrigidas do efeito da radiação radial	8
(2) Coordenadas imagem corrigidas dos efeitos das distorções radial e tangencial	8
(3) Equações de colinearidade	10
(4) Razão dupla.....	11
(5) Razão dupla pela lei dos senos	11
(6) Razão dupla.....	12
(7) Transformação dos oito parâmetros	12
(8) Gradiente da imagem	29
(9) Percentagem de Detecção de Fachadas.....	35
(10) Percentagem de Qualidade	35
(11) Percentagem de Falsos Negativos	35

Lista de acrónimos e abreviaturas

3D – tridimensional

ANAC - Autoridade Nacional da Aviação Civil

API - Application Programming Interface

DEM - Digital Elevation Model

DN – Digital Number

DSM – Digital Surface Model

DTM – Digital Terrain Model

EMQ - Erro médio quadrático

ETRS89/PT-TM06 - European Terrestrial Reference System 1989/ Portugal – Transversa de Mercator 2006

IMU - Inertial Navigation Unit

INS - Inertial Navigation System

GNSS – Global Navigation Satellite System

GPS - Global Positioning System

GSD - Ground Sample Distance

LiDAR - Light Detection And Ranging

OE – Orientação externa

OI – Orientação interna

PDI - Processamento Digital de Imagem

PF - Ponto Fotogramétrico

RGB – Red, Green, Blue

SfM – Structure from Motion

TIN - Triangulated Irregular Network

UAV - Unmanned Aerial Vehicle

WGS84 - World Geodetic System 1984

1 Introdução

1.1 Enquadramento

No seguimento da evolução tecnológica têm vindo a surgir novas aplicações baseadas na disponibilização de modelos tridimensionais (3D), localizados em ambientes urbanos, e que proporcionam aos utilizadores um ambiente virtual georreferenciado, que pode ser utilizado em várias áreas, tais como o planeamento urbano, turismo, cultura, gestão paisagística, manutenção de infraestruturas, imobiliário, entre outras. Os cenários urbanos 3D são assim uma das áreas que mais desafios coloca aos sistemas de visualização interativa (Fonseca, 2008).

Muitas tecnologias têm vindo a ser desenvolvidas para melhorar a captura da realidade, tanto através de novo *software*, como de novos equipamentos. Atualmente, as técnicas mais utilizadas para obter a geometria de uma área de uma cidade, a grande escala, de forma automática, ou semiautomática, são o *laser scanner* aéreo (LiDAR – ‘*Light Detection And Ranging*’) e a fotogrametria aérea. Por vezes, as duas técnicas são até utilizadas em conjunto, de forma a obter modelos com maior precisão e realismo, uma vez que a primeira apresenta melhor precisão, e a segunda captura texturas realistas. Contudo, com os novos *software* desenvolvidos para o processamento de imagens através do método *Structure from Motion* (SfM), já é possível obter nuvens com precisão equivalente à obtida com LiDAR, a partir de coberturas fotográficas.

A texturização é um dos principais pontos de atenção de um modelo 3D, uma vez que cria a falsa impressão de um nível de detalhe geométrico mais elevado. É possível obter modelos 3D com um elevado nível de detalhe geométrico nas fachadas e com textura realista usando imagens captadas a partir de um automóvel que vai mapeando a cidade, mas isto requer um processo extra se o modelo inicial foi obtido a partir de processos de fotogrametria aérea (Smith *et al.*, 2009).

A vulgarização dos modelos 3D e de outros produtos da fotogrametria, que até há muito pouco tempo, estavam diretamente conotados com elevados custos, tanto económicos como de processamento, são uma realidade inquestionável devido à excecional evolução tecnológica das câmaras digitais, ao poder de cálculo e armazenamento dos computadores pessoais e à acessibilidade generalizada aos UAV, que são uma ferramenta de aquisição de dados aerofotogramétricos bastante mais económica que outras tecnologias existentes, e que tem vindo cada vez mais a impor a sua presença em alguns mercados.

Uma característica das imagens aéreas é a presença de distorção radial, causada pelas elevações do terreno e objetos nele presentes, permitindo visualizar algumas fachadas dos edifícios que se encontram afastados do nadir da imagem, ao mesmo tempo que, através de técnicas de fotogrametria, se obtém o modelo 3D. Em imagens obtidas com UAV, estas distorções são ainda mais pronunciadas, relativamente à fotogrametria aérea tradicional, devido à baixa altitude de voo e à instabilidade do veículo durante o mesmo, pelo que estes poderão ser uma boa alternativa, e bastante mais económica, à fotografia terrestre para aquisição de fachadas, em zonas em que as condições o permitam.

1.2 Estado da Arte

Várias abordagens têm sido desenvolvidas para a extração de texturas de fachadas de edifícios. O grande desafio é automatizar o máximo possível este processo.

Rau *et al.* (2012) desenvolveram um algoritmo que, a partir de imagens aéreas oblíquas e da sua orientação externa e interna, verifica para cada fachada de um modelo 3D previamente elaborado, e para

cada uma das imagens, a visibilidade da fachada nas mesmas. De seguida, faz uma análise às imagens para detetar oclusões nas fachadas e combina informação de várias imagens para obter texturas livres de oclusões.

Frueh *et al.* (2004) desenvolveram um método com o qual, a partir de imagens aéreas oblíquas e de um modelo 3D obtido com laser scanning, as linhas 2D nas imagens são projectadas no modelo 3D e, para cada triângulo no modelo, a imagem óptima é seleccionada tendo em conta a existência de oclusões, a resolução da imagem, a coerência com triângulos vizinhos e a orientação da normal ao plano da fachada.

Em Zebedin *et al.* (2007) é apresentada uma abordagem interessante. Trata-se de um algoritmo para estimar a posição precisa dos planos das fachadas em DSMs (*Digital Surface Models*) reconstruídos a partir de imagens aéreas. Assume-se que a fachada é um plano vertical e que o DSM é preciso o suficiente para gerar uma hipótese inicial para a posição da fachada, que depois irá sofrer um processo de otimização. A inicialização é primeiro orientada grosseiramente segundo as direções da linha principal de textura. Depois, um algoritmo hierárquico realiza uma otimização mais fina, a cada iteração, para maximizar a correlação entre diferentes vistas da mesma fachada. A cada iteração, duas vistas do mesmo edifício são comparadas e a orientação da fachada vai sendo alinhada até ficar paralela à fachada real. Depois, é calculado o histograma de orientações da imagem. Espera-se encontrar, neste histograma, quatro picos próximos dos ângulos 0°, 90°, 180° e 360°, que representam os *pixels* situados nas arestas das fachadas e janelas. É aplicada uma correção à orientação da fachada, de forma a que os picos do histograma correspondam exatamente aos ângulos pretendidos. Por último, o algoritmo aplica uma abordagem em pirâmide, na qual, se vai aumentando a resolução da imagem, a cada iteração, e é determinada a translação sofrida pelo limite da fachada, até o vetor translação encontrado for nulo e já não ser possível aumentar a resolução da imagem.

1.3 Objetivos e contribuição do trabalho

O objetivo principal deste trabalho é criar um algoritmo para gerar, automaticamente, um atlas de texturas das fachadas de edifícios, consoante a morada e número de polícia, para uma determinada área, a partir de imagens obtidas com UAV.

Os objetivos específicos deste trabalho são:

- Elaborar uma ortofoto e DEM a partir das imagens obtidas com UAV;
- Desenvolver um algoritmo que identifique automaticamente edifícios presentes na área coberta pelas imagens, consoante a morada e número de polícia;
- Desenvolver um algoritmo para reconhecimento e retificação das fachadas;
- Interligar os dois algoritmos de forma a obter um método capaz de extrair fachadas, consoante a morada e número de polícia;
- Avaliar a qualidade do método.

Pretende-se, com este algoritmo, minimizar o tempo necessário para a texturização de modelos 3D, ao mesmo tempo que se apresenta uma alternativa mais económica em relação aos métodos utilizados, que passam pela aquisição de fotografias em campo.

1.4 Organização do Trabalho

Este documento encontra-se dividido em 6 capítulos. O presente capítulo apresenta um enquadramento, o estado da arte referente ao tema desta dissertação e os objetivos e o contributo do mesmo.

No segundo capítulo é apresentada a fundamentação teórica por detrás dos dados e métodos utilizados neste trabalho e o terceiro capítulo contém uma descrição dos dados primários utilizados (imagens e respetivas orientações) e sua aquisição.

O quarto capítulo é referente às operações preliminares efetuadas no *software* PhotoScan, com o intuito de obter um DEM e ortofoto que farão parte dos dados *input* do algoritmo apresentado neste documento.

O quinto capítulo apresenta o algoritmo realizado em linguagem de programação MATLAB e uma análise ao mesmo.

Por último, no sexto capítulo apresenta-se as conclusões.

2 Fundamentos teóricos

2.1 UAV

A sigla UAV vem da língua inglesa ‘*Unmanned Aerial Vehicle*’, que em português significa Veículo Aéreo não Tripulado, VANT. Os UAV são então veículos aéreos não tripulados, ou seja, não existe qualquer presença humana a bordo da aeronave, e podem ser controlados tanto numa estação em terra, como por outra aeronave, ou podem ser programados para voar autonomamente. Estes veículos são ainda denominados de *drones*, designação originária dos Estados Unidos da América, e que significa “zangão”, na tradução literal para a língua portuguesa.

O conceito UAV não se resume simplesmente à plataforma, podendo ser visto como um sistema que abrange o próprio veículo, *hardware* e *software*, estação de controlo, sensores e equipamentos de navegação (Roberto, 2013).

Os UAV foram desenvolvidos para fins militares, com objetivos de espionagem, vigilância, reconhecimento e mapeamento de áreas hostis (Roberto A. J., 2013), tendo começado a aparecer durante a Primeira Guerra Mundial. O desenvolvimento desta tecnologia para aplicações civis foi surgindo mais lentamente, principalmente devido aos custos associados à sua construção.

Como qualquer sistema, este apresentava algumas inconsistências, principalmente em manter a altitude prevista e a navegação da própria aeronave era muito instável, condicionando, em algumas situações, o pós-processamento dos dados recolhidos pelos sensores, mas, com o desenvolvimento tecnológico nesta área, foram surgindo novos sistemas que permitiram a integração dos Sistemas GPS/INS, ‘*Global Positioning System/ Inertial Navigation System*’, em novos modelos de UAV, melhorando a estabilidade do voo.

Com o passar do tempo, devido aos avanços tecnológicos, a sua aplicação pôde ser ampliada para o uso civil. No período de 2004 a 2007 foram desenvolvidos projetos em que foram usados UAV com asas rotativas, como é o caso do *Copeter 1B*, desenvolvido pela empresa *Surveycopter*. Uma das características principais era a capacidade de efetuar voo de forma autónoma, precisamente, seguindo o percurso estabelecido (Roberto, 2013).

Hoje em dia o termo UAV é muito usado nas ciências da computação, robótica e inteligência artificial, bem como na fotogrametria e deteção remota, sendo por isso um conceito vulgar, mas bastante promissor, pois várias novas aplicações têm sido desenvolvidas nos últimos anos.

2.1.1 Uso de UAV na Fotogrametria

Atualmente, com a ajuda dos UAV e da comunidade que desenvolve ferramentas de processamento dos dados, existe a possibilidade de tornar a realização de produtos da fotogrametria (ortofotos, DSM, etc.) uma tarefa mais agilizada e utilizá-la num maior leque de aplicações, o que, num passado recente não faria sentido devido aos seus elevados custos e à sua morosidade. A capacidade de intervenções rápidas e/ou sistemáticas, tornam os UAV a vanguarda de muitas operações, como por exemplo a análise de cenários de catástrofe ou a agricultura de precisão.

Para além do fator económico, os UAV têm também a vantagem do fator tempo em projetos em que as alterações ao objeto são consideravelmente rápidas e onde o recurso a plataformas tradicionais seria incomportável, como por exemplo em zonas de implantação de grandes obras de engenharia.

As principais diferenças entre a obtenção de imagens por plataformas tradicionais e por UAV são a altura de voo e a atitude da aeronave. Voos com UAV são realizados com altitude mais baixa, tornando necessária a captação de mais imagens para que exista sobreposição suficiente para gerar estereoscopia. Os UAV são veículos mais leves e por isso mais instáveis, tornando necessário ter em consideração, durante o planeamento, que a sobreposição entre imagens deverá ser maior que nos voos tradicionais, para evitar lacunas. O fator instabilidade da plataforma faz com que existam elevadas distorções de perspectiva nas imagens obtidas com UAV, o que no contexto deste trabalho é uma vantagem, pois permite uma melhor visualização das fachadas.

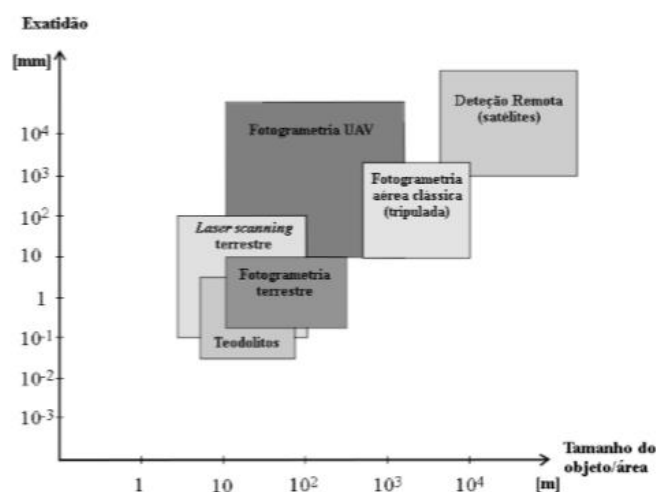


Figura 1- Comparação de alguns métodos de medição no que diz respeito à exatidão e tamanho do objeto/área. (Fonte: Serra, 2013)

A figura 1 apresenta uma comparação da exatidão e área coberta por alguns métodos. Este método, Fotogrametria com UAV, apresenta uma exatidão comparável a outros métodos aéreos, podendo mesmo substituí-los em áreas não muito extensas. Uma das limitações dos UAV é a duração da bateria, que não permite voos muito longos.

2.1.2 Viabilidade legal do uso de UAV

A questão da segurança e regulamentação dos UAV tem sido fortemente posta em causa. A operação massiva e desregulada dos mesmos pode afetar a segurança de pessoas e bens, bem como a segurança da própria operação. Apesar de ainda não existir, a nível internacional ou europeu, legislação harmonizada específica para a utilização de UAV, foram criadas, recentemente, normas nacionais que determinam as condições aplicáveis à operação e utilização destas aeronaves no espaço aéreo nacional. Estas normas foram colocadas para consulta pública em Maio de 2016 pela Autoridade Nacional da Aviação Civil (ANAC) a quem compete a criação destas normas (ANAC, 2016).

Conforme estipulado no artigo 3º, apenas se podem efetuar voos até 120 metros de altura; a operação deve ser executada de forma a minimizar riscos para as pessoas, bens e outras aeronaves; e é possível realizar voos, desde que com solicitação e autorização prévia pela ANAC, exceto em zonas de tráfego de aeródromo e outras áreas referidas no artigo 5º do documento.

Não existem normas específicas para os voos realizados em zonas urbanas. Porém, devido às limitações em relação à altura de voo, o uso de UAV poderá não ser uma opção viável em áreas urbanas com edifícios muito altos. Se a área urbana se encontrar próxima de um aeródromo, o voo poderá não ser permitido.

2.2 Geometria da fotografia e projeção central

2.2.1 Imagem digital

Até à data, a maioria dos dispositivos de captação de imagens aéreas fornecem produtos digitais, uma vez que a detecção é normalmente realizada pela conversão da radiação eletromagnética recebida, num sinal elétrico de saída que é subsequentemente digitalizado em números digitais (DNs). Cada fotodetector produz um pixel e uma imagem é um conjunto de *pixels* que pode ser matematicamente representada por uma matriz $M \times N$ de números, sendo que M e N representam as dimensões da imagem (Verhoeven *et al.*, 2013) e o valor de cada pixel representa uma função da quantidade física de radiação eletromagnética medida.

O mesmo pixel pode ainda possuir vários DNs, correspondentes às radiações medidas em bandas espectrais diferentes. Neste caso a imagem será representada por um conjunto de matrizes. As imagens coloridas, mais vulgarmente utilizadas, captam radiação em três bandas espectrais (vermelho, verde e azul), daí o termo ‘imagem RGB’, RGB do inglês *red*, *green* e *blue*.

Cada imagem é ainda caracterizada pelo número de *bits*, que determina a resolução através da qual as amplitudes do sinal de radiação contínua podem ser mapeadas num conjunto discreto de valores digitais (Verhoeven *et al.*, 2013). Por outras palavras, quanto maior o número de *bits*, maior é a escala de cores disponível.

No entanto, a radiação eletromagnética que chega ao sensor da imagem é degradada de várias formas, portanto a imagem digital final nunca será uma reprodução 100% fiel da cena do mundo real. Fatores como a curvatura terrestre, plano imagem não plano, efeitos de refração atmosférica, distorções óticas e a inclinação da câmara e variações no terreno provocam distorções geométricas nas imagens.

Na fotogrametria e na visão computacional, a geometria da projeção central, ou perspectiva central, é usada para modelar matematicamente os elementos de uma imagem. Na fotogrametria, isso é expresso pelas equações de colinearidade, descritas no capítulo 2.2.5, que indicam que o ponto objeto, o centro de projeção da câmara e o ponto imagem se encontram sobre uma linha reta e a imagem é formada sobre um plano exato (Verhoeven *et al.*, 2013). Contudo, fatores como as distorções da lente (radial e tangencial), descritas no capítulo 2.3.3, efeitos atmosféricos (refração) e um sensor imagem não-plano, impedem que os três pontos referidos se encontrem exatamente sobre uma linha reta e que a imagem se forme sobre um plano exato. Como o efeito da refração atmosférica apenas tem relevância quando os raios luminosos percorrem grandes distâncias, por exemplo, em imagens captadas a elevadas altitudes ou com ângulos oblíquos, apenas se irão considerar, neste documento, os efeitos das distorções das lentes.

2.2.2 Parâmetros que caracterizam a geometria da imagem

Os parâmetros matemáticos que descrevem a situação ideal, na qual a projeção central é perfeita, transformando objetos 3D em objetos 2D numa imagem, sem influência das distorções causadas pelas lentes da câmara são:

- Distância focal (c): distância medida ao longo do eixo ótico, desde o centro ótico do sistema de lentes (ou nodo posterior em lentes compostas) até ao plano da imagem (plano focal);
- Coordenadas do ponto principal (x_0 , y_0): corresponde à interseção do eixo ótico do sistema de lentes com o plano da imagem. As coordenadas do ponto principal variam com as definições de

zoom, mas estarão sempre próximas do ponto central da imagem. Numa câmara ideal, o ponto principal deverá coincidir com o ponto central da imagem;

- Parâmetros de distorção radial (k_1, k_2, k_3, k_4): Coeficientes de um polinómio que corrige a imagem da distorção radial causada pelo sistema de lentes da câmara. A distorção radial causa distorções ao longo de linhas radiais a partir do ponto principal de simetria;
- Parâmetros de distorção tangencial (p_1, p_2): Parâmetros de equações que corrigem a imagem da distorção tangencial, que corresponde à distorção assimétrica do sistema de lentes da câmara. Ambas as distorções tangencial e radial são causadas por imperfeições na construção e alinhamento das lentes.

Os parâmetros acima descritos constituem os **parâmetros de orientação interna da câmara**. As correções a efetuar às coordenadas imagem de um ponto, de forma a evitar a influência das distorções radial e tangencial são descritas em 2.2.3.

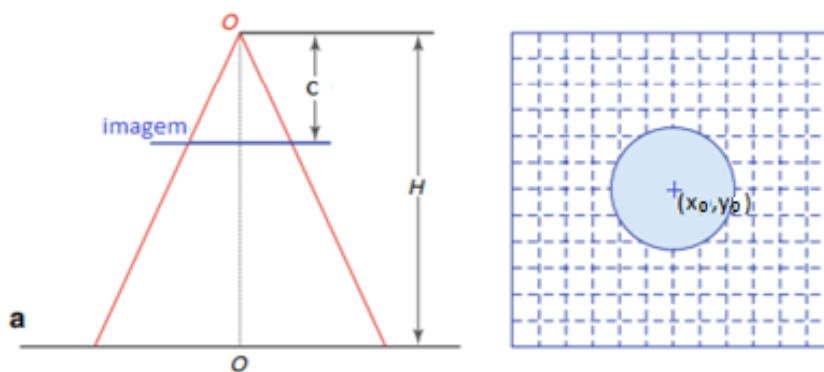


Figura 2- Perspectiva central de uma imagem obtida na vertical e alguns dos parâmetros de orientação interna (distância focal, c , e coordenadas do ponto principal (x_0, y_0)). (Adaptado de Verhoeven *et al.*, 2013)

No momento da tomada de foto, a câmara encontra-se numa certa posição no espaço, e com uma certa orientação. Esta localização é definida pelas coordenadas terreno do centro de projeção (X_0, Y_0, Z_0) e a orientação é definida pelos ângulos (ω, ϕ, κ) que descrevem as rotações do sistema de coordenadas da câmara em relação ao sistema de coordenadas terreno. Estes constituem os **parâmetros de orientação externa** de uma imagem.

2.2.3 Correção das coordenadas imagem

Como se refere em 2.2.2, para além da distância focal, c , e das coordenadas do ponto principal, x_0 e y_0 , fazem parte da orientação interna das imagens as funções que descrevem a distorção radial e a distorção tangencial. A primeira é a distorção causada pelo sistema de lentes da objetiva e a segunda é a distorção causada pela descentragem das diversas lentes. Estas distorções são geralmente determinadas em laboratório, podendo ser analiticamente reajustadas após o voo. Este reajustamento justifica-se quando existem grandes diferenças entre as condições de pressão e temperatura no laboratório e no ambiente de voo (Redweik, 2010).

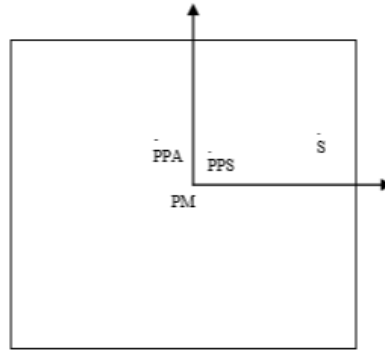


Figura 3- Origens dos diferentes sistemas de coordenadas foto (Fonte: Redweik, 2011)

Sendo S um ponto com coordenadas foto (x,y) referentes ao ponto principal de simetria (PPS), que é o centro médio dos círculos de igual distorção radial, a equação 1 apresenta as coordenadas de S corrigidas da distorção radial, obtendo-se as coordenadas em relação ao ponto médio da fotografia (PM).

$$\begin{cases} x' = x - (k_1d^2 + k_2d^4 + k_3d^6) \times \frac{y}{d} \\ y' = y - (k_1d^2 + k_2d^4 + k_3d^6) \times \frac{x}{d} \end{cases} \quad (1)$$

sendo d a distância ao centro, dada por $\sqrt{x^2 + y^2}$ e k1, k2 e k3 parâmetros da distorção radial.

A equação 2 apresenta as coordenadas corrigidas também da distorção tangencial, e referentes ao ponto de auto-colimação (PPA).

$$\begin{cases} x'' = x' - p_1(d^2 + 2x'^2) + 2p_2x'y' \\ y'' = y' - p_2(d^2 + 2y'^2) + 2p_1x'y' \end{cases} \quad (2)$$

sendo p1 e p2 parâmetros da distorção tangencial.

2.2.4 Efeitos do relevo numa imagem

Qualquer imagem, mesmo que obtida exatamente na vertical, possui desvios causados pelo relevo do terreno e diferenças de altura de objetos nele presentes. Estes desvios são chamados de distorção radial causada pelo relevo e altura de objetos. Assim, qualquer ponto no cenário fotografado situado acima ou abaixo de uma superfície de referência horizontal, que indica a elevação do terreno no nadir fotográfico, irá sofrer um desvio devido à perspectiva central da fotografia aérea.

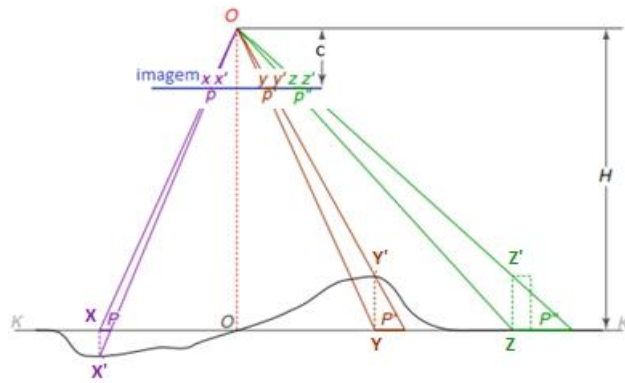


Figura 4- Distorção radial causada pela topografia e altura dos objetos no terreno. (Adaptado de Verhoeven *et al.*, 2013)

A figura 4 mostra a aquisição de uma fotografia exatamente na vertical. KK' representa o plano de referência. Numa projeção ortogonal, o objeto representado à direita teria as mesmas coordenadas imagem para o ponto situado na base, Z , e para o ponto situado no topo, Z' . No entanto, devido à projeção central, o topo é representado em z'' , em vez de z , sofrendo um desvio de magnitude p'' , e o lado esquerdo do objeto é visível na imagem (Verhoeven *et al.*, 2013). No caso de X , que se encontra abaixo do plano de referência, o desvio sofrido irá aproximar o correspondente ponto imagem, x' , do centro da imagem, em vez de o afastar.

O ponto localizado no nadir não sofre deste desvio, pois a projeção nesse ponto é ortogonal. À medida que nos afastamos do centro da imagem, os desvios aumentam e por isso representam uma distorção radial.

Para corrigir estas deformações, há que transformar a projeção central numa projeção ortogonal para o plano de referência horizontal e remover a inclinação da fotografia, caso exista. A este processo dá-se o nome de ortorretificação.

Embora este fenómeno de distorção complique o mapeamento e a interpretação da imagem aérea, permite aos seres humanos ter perceção das três dimensões e calcular a altura dos objetos a partir das imagens, que é exatamente o que se pretende neste trabalho, ou seja, tirar partido da distorção radial de edifícios que se encontrem na periferia das imagens para extrair as suas fachadas.

2.2.5 Equações de colinearidade

Como já foi referido anteriormente, admite-se que no momento em que a fotografia foi captada, existia colinearidade entre o centro de projeção, o ponto objeto e o seu correspondente ponto imagem e portanto um ponto P no espaço conecta-se ao seu ponto imagem correspondente, P' , através de uma linha reta que passa pelo centro de projeção O (figura 5). As equações de colinearidade (equação 3) são as equações fundamentais da fotogrametria pois relacionam coordenadas foto $(x, y, 0)$ e coordenadas objeto (X, Y, Z) do mesmo ponto, na altura em que a fotografia foi tirada.

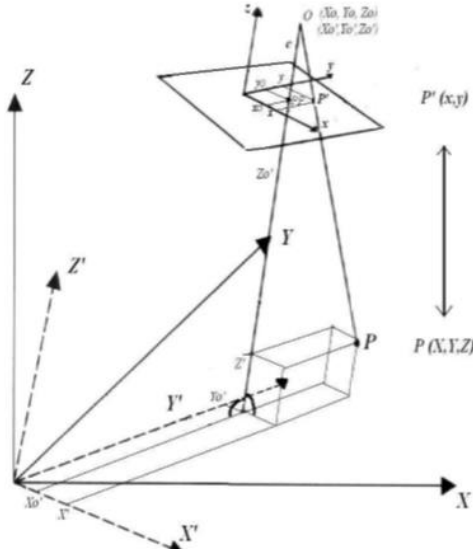


Figura 5- Relação entre imagem e objeto (Redweik, 2010)

$$\begin{cases} x = x_0 - c \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \\ y = y_0 - c \frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \end{cases} \quad (3)$$

em que (x_0, y_0, c) são as coordenadas foto do centro de projeção, c é a constante da câmara ou distância focal, (X_0, Y_0, Z_0) são as coordenadas do centro de projeção e $r_{i,j}$ os elementos da matriz de rotação espacial entre os sistemas de coordenadas foto e objeto. Esta matriz de rotação é função dos ângulos ómega (ω), fi (φ) e K, parâmetros de orientação externa da imagem. É possível consultar a dedução destas equações em Redweik (2010).

2.2.6 Transformação dos 8 parâmetros

Tal como já se referiu anteriormente, sendo a fotografia uma projeção central do objeto fotografado sobre o plano imagem, o modelo geométrico adotado pela Fotogrametria para o processamento métrico de uma fotografia é a perspectiva central, que é uma transformação geométrica de um objeto sobre um plano e que tem várias propriedades apresentadas de seguida, as quais podem ser consultadas em Redweik (2008):

Qualquer ponto do objeto é transformado num ponto na imagem: Qualquer ponto P no objeto e o respetivo ponto P' na imagem estão sobre uma reta designada por raio projetivo. O conjunto dos raios projetivos dos diversos pontos objeto denomina-se feixe perspetivo e intersesta-se num único ponto, denominado centro de projeção ou pólo da perspetiva.

A figura 6 apresenta uma representação de uma projeção central de um objeto na imagem.

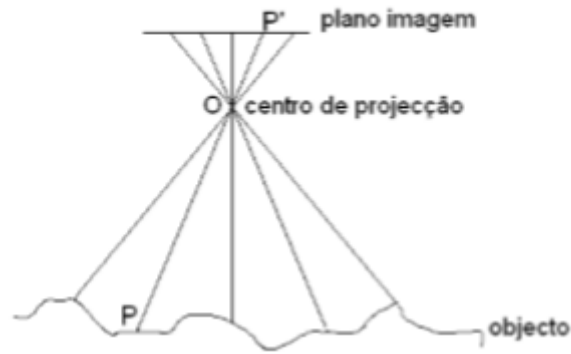


Figura 6- Projeção central de um objeto 3D num plano (Fonte: Redweik, 2008)

Qualquer reta do objeto é transformada numa reta na imagem, à exceção das retas que passem pelo centro de projeção: As retas que passam pelo centro de projeção são transformadas em pontos. Entre a reta objeto e a reta imagem é válida a mesma razão dupla entre segmentos de reta definidos por quatro pontos correspondentes numa e noutra reta. A figura 7 ilustra a projeção perspectiva de uma reta e a equação 4 exprime a razão dupla entre os segmentos de reta definidos pelos pontos DEFG.

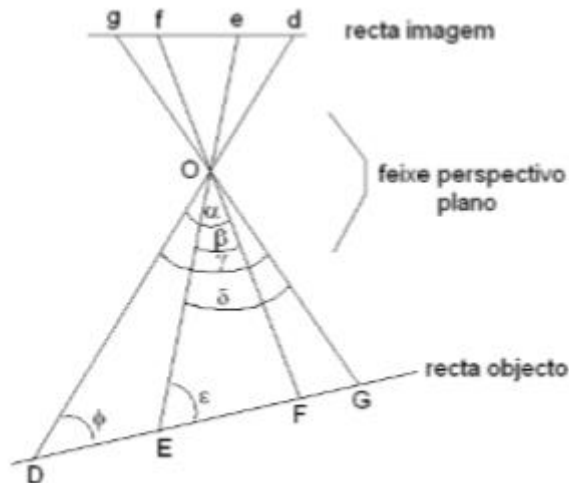


Figura 7- Projeção perspectiva de uma reta (Fonte: Redweik, 2008)

$$r = \frac{DF}{DG} \times \frac{EG}{EF} \quad (4)$$

A razão dupla é invariante na projeção central entre retas: Aplicando a lei dos senos aos triângulos existentes e tendo em conta que os ângulos de um lado e do outro do centro de projeção são iguais, vem que:

$$r = \frac{\sin \alpha}{\sin \gamma} \times \frac{\sin \delta}{\sin \beta} \quad (5)$$

Verifica-se que o valor da razão dupla depende apenas dos ângulos que os raios projetivos fazem no centro de projeção, sendo por isso igual para qualquer reta que atravessasse o feixe, independentemente da sua inclinação.

$$r = \frac{DF}{DG} \times \frac{EG}{EF} = \frac{df}{dg} \times \frac{eg}{ef} \quad (6)$$

Com esta propriedade é possível, com um mínimo de três pontos conhecidos nas duas retas, determinar onde se localiza um ponto imagem, apenas conhecendo um dos segmentos que o seu ponto objeto define.

A invariância da razão dupla entre retas é também válida para planos: Considerem-se 5 pontos projetados perspetivamente de um plano sobre outro.

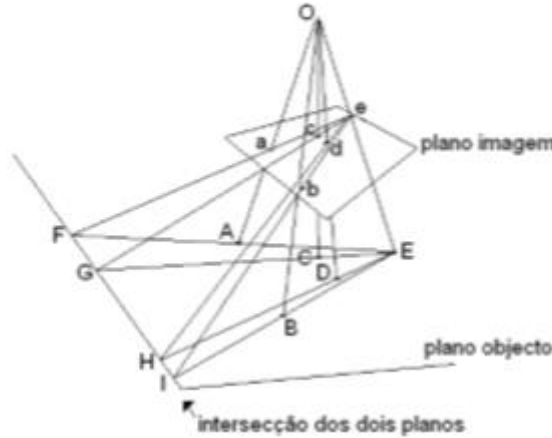


Figura 8- Perspetiva entre planos (Fonte: Redweik, 2008)

Se tomarmos um dos 5 pontos como pólo de um feixe perspetivo plano no plano objeto, cujos raios passam pelos outros 4 pontos, vimos que a razão dupla dos segmentos definidos por esses pontos ao longo desse feixe é invariante. No plano imagem, também se pode definir um feixe perspetivo plano com os raios definidos pelos 4 pontos imagem dos anteriores. Acontece que os raios perspetivos dos dois feixes planos se intersectam nos pontos sobre a reta de intersecção dos dois planos. Assim sendo, a razão dupla que é válida para um dos feixes é válida para o outro. A razão dupla entre planos fica assim definida por 4 pontos, entre os quais nenhum trio seja colinear.

A transformação perspetiva entre planos, também denominada por transformação dos oito parâmetros relaciona coordenadas de um ponto num plano com as coordenadas da sua imagem perspetiva no outro plano:

$$\begin{cases} X = \frac{e1x + f1y + g1}{e0x + f0y + 1} \\ Y = \frac{e2x + f2y + g2}{e0x + f0y + 1} \end{cases} \quad (7)$$

Com 4 pontos idênticos (conhecidos nos dois planos) formulam-se 8 equações das quais se podem calcular os 8 parâmetros da transformação. Conhecendo estes parâmetros, é então possível transformar qualquer ponto de um plano para o outro.

2.3 Aquisição e processamento de dados com UAV

Avanços no domínio da Visão Computacional, nomeadamente a criação de novos algoritmos para processamento de fotografias, têm permitido alcançar um nível de automatização muito mais elevado e facilitar a utilização de *software* fotogramétrico. É exemplo destes avanços a abordagem *Structure from*

Motion (SfM), que recorre ao *Scale Invariant Feature Transform* (SIFT) para criar modelos 3D a partir de uma série de imagens com sobreposição (Turner *et al.*, 2012).

O método SfM difere da fotogrametria convencional, fundamentalmente pelo fato de a geometria da cena, as posições da câmara e a orientação serem resolvidas automaticamente sem a necessidade de especificar *a priori* as posições das câmaras, ou uma rede de pontos com coordenadas conhecidas. Em vez disso, estes parâmetros são resolvidos simultaneamente através de um ajustamento iterativo, baseado num conjunto de dados extraídos automaticamente de uma série de imagens com sobreposição (Westoby *et al.*, 2012).

Esta técnica é ideal para processar imagens obtidas com UAV pois é mais eficaz em pequenas áreas de estudo, onde as imagens têm resolução suficiente para capturar textura detalhada (Fonstad *et al.*, 2013). Para além disso, ao contrário do *software* fotogramétrico tradicional, o método é robusto em situações em que as imagens possuem grandes mudanças de orientação, escala e translações entre imagens. É possível obter nuvens de pontos com qualidade comparável à obtida com LiDAR, mas com uma facilidade de utilização inédita e um custo muito reduzido.

A figura 9 apresenta os passos para a obtenção de produtos fotogramétricos a partir de imagens obtidas com UAV, desde a aquisição de imagens, até à obtenção do modelo 3D e ortofoto.

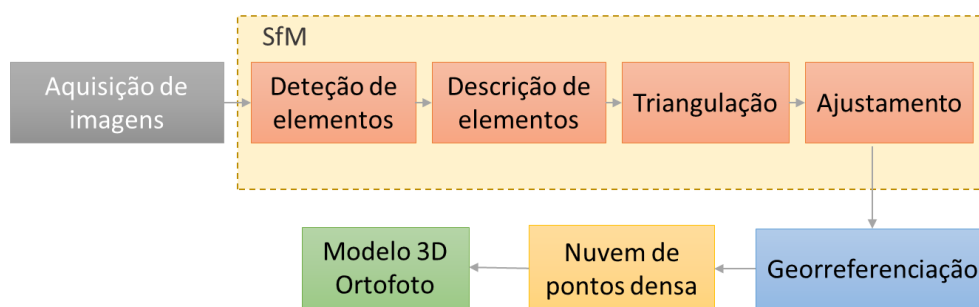


Figura 9- Passos do SfM+MVS. (Adaptado de Verhoeven *et al.*, 2013)

2.3.1 Aquisição de imagens e medição de PFs

Após o planeamento do voo, durante o qual se define a área geográfica a levantar, altura de voo, a resolução espacial desejada, ou GSD, do inglês ‘*Ground Sample Distance*’, o número de fiadas e coordenadas dos pontos de tomada de foto, é então realizado o voo.

Para além da câmara, a aeronave vai equipada com *Global Navigation Satellite System* (GNSS) e *Inertial Measuring Unit* (IMU). A sua presença a bordo é essencial para o controlo da rota da aeronave. O GNSS fornece a posição absoluta do sensor e o IMU fornece as acelerações lineares e angulares em cada momento de voo. Este é reiniciado a cada medição com GNSS. Quando o seu sinal falha, é possível obter as posições absolutas e atitude através das medições do IMU. A sua presença, juntamente com a presença do piloto automático, permite ao UAV cumprir a missão autonomamente.

Durante o voo, a aeronave é vigiada por uma estação de controlo terrestre onde é possível verificar a posição, velocidade, atitude, observações GNSS e estado da bateria, em tempo real.

Para que o algoritmo SfM funcione bem, um ponto em duas imagens sucessivas não deve mudar o ponto de vista em mais de 25°-30°, pelo que se torna importante garantir sobreposições de 60%-80%.

Visto que as coordenadas obtidas para os pontos de tomada de foto com o sistema GNSS/IMU são pouco precisas, é necessário levantar um conjunto de pontos fotogramétricos (PFs) para mais tarde georreferenciar o bloco.

Os PFs são pontos identificáveis nas fotografias e que são passíveis de ser medidos no terreno. A precisão exigida para a determinação das suas coordenadas depende da precisão que se pretende obter no produto final, mas sempre que possível são medidos por métodos de coordenação por GPS.

A localização dos PFs deve ser feita com base no mapa de voo da zona a levantar. Devem estar bem distribuídos pelo bloco e apresentar uma localização estratégica, evitando a proximidade a objetos altos, de forma a estarem bem visíveis nas imagens. O mínimo teórico de PFs exigido para orientar absolutamente o bloco são 3 PFs. Contudo, devem-se utilizar mais PFs para obter redundância. A distribuição dos PFs também é um aspeto importante. Estes devem-se encontrar bem espalhados pelo bloco, com especial foco nas extremidades do bloco.

2.3.2 Técnica SfM

Após a aquisição de dados e sua introdução no *software*, é então aplicada a técnica SfM, a qual se divide em quatro etapas, como é possível observar na figura 9.

Existem várias abordagens de **deteção de elementos** que variam na eficácia, complexidade computacional e no tipo de elementos detetados (arestas, regiões de interesse ou pontos de interesse). A abordagem mais aplicada na SfM compreende a deteção de pontos de interesse (pontos de correspondência) (Verhoeven *et al.*, 2013). Os pontos de interesse são pontos na imagem facilmente identificáveis por estarem cercados por texturas diferentes e devem ser visíveis independentemente das condições de iluminação, ruído na imagem, escala, etc. O SIFT é um dos algoritmos mais populares para o cálculo de pontos de interesse. Baseia-se num processo de correspondência automática de imagens, utilizando imagens posicionadas aleatoriamente e que vai procedendo ao reconhecimento de elementos nas imagens invariantes à escala e à rotação e parcialmente invariáveis às condições de luminosidade e à perspetiva da câmara. Em blocos com muitas imagens, o algoritmo poderá perder muito tempo a realizar esta tarefa, pois irá verificar a correspondência entre todas as imagens, mesmo as que não têm pontos em comum. Por este motivo, a introdução prévia das coordenadas dos pontos de tomada de foto obtidos pelo sistema GNSS/IMU poderá agilizar esta tarefa, pois indica ao *software* qual a sequência das imagens.

Na tarefa **descrição de elementos** são procuradas correspondências entre pontos de interesse de imagens com sobreposição através de características descritoras dos elementos que se assemelham. A figura 10 a) mostra várias imagens de um cubo obtidas com perspetivas diferentes e a correspondência obtida entre pontos de interesse (pontos assinalados com a mesma cor representam o mesmo ponto 3D no objeto). A figura 10 b) apresenta uma representação do objeto e das imagens com as respetivas posições e orientações associadas.

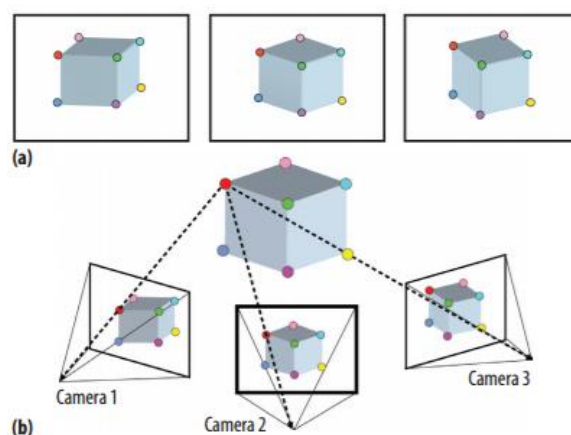


Figura 10- Imagens de um cubo com perspectivas diferentes (a) e reconstrução do cubo (b) (Fonte: Agarwal, 2010)

A correspondência entre pontos de interesse é realizada através dos chamados descritores ou vetores de elementos, que calculam as características locais que descrevem a vizinhança em torno de cada ponto de interesse. Tal como o ponto de interesse, este vetor deve ser invariante, ou seja, deve ser robusto o suficiente para ser imune a deslocamentos, ruído na imagem e deformações geométricas e fotogramétricas nas mesmas (Verhoeven *et al.*, 2013). A detecção de características locais é mais adequada ao zoom e orientação da imagem, ruído, etc. do que a detecção de características globais, que se baseiam na área, perímetro, ou simetria de um objeto, pois mais facilmente, estas últimas, são afetadas pelos fatores referidos.

Existem vários métodos para descrever a vizinhança de um ponto de interesse. O SIFT é exemplo de um desses métodos e divide-se em quatro estágios: Detecção de extremidades (efetua uma pesquisa sobre todas as escalas locais da imagem, pelos pontos de interesse), Localização de pontos chave (os pontos detetados na etapa anterior são classificados como pontos-chave com base na sua estabilidade), Atribuição de orientação (a orientação é atribuída a cada ponto chave com base na direção do gradiente, tornando o descritor invariante para orientação, escala e transformações locais) e, por último, a Descrição de pontos-chave (são medidos gradientes dentro de uma pequena janela em torno do ponto central. Estes gradientes são resumidos num vetor descritor, que define uma tolerância às alterações de iluminação e distorção da forma local) (Lowe, 2004).

Uma vez detetados os elementos numa imagem, estes podem ser combinados em pares de imagens diferentes através do encontro de elementos semelhantes (Agarwal *et al.*, 2010). Os pares de correspondências são interligados a partir de pontos dos elementos, em conjunto, para formar linhas correspondentes ao mesmo ponto 3D na cena. Para calcular a correspondência é utilizada a distância entre os descritores. O tamanho do descritor tem influência no tempo de execução: quanto menor for o descritor, mais rápida será a correspondência de pontos de interesse, mas geralmente é menos distinguível. Complementarmente, existem algoritmos que asseguram a rejeição da probabilidade de falsa correspondência, testando a sua consistência. Este processo é realizado em todos os pares possíveis, verificando se as correspondências cumprem a restrição da geometria epipolar (Verhoeven *et al.*, 2013). No final deste processo são obtidas as matrizes fundamentais 3×3 dependentes dos sete parâmetros que descrevem o movimento entre duas imagens consecutivas (três ângulos, três translações e um fator de escala).

O conjunto de pontos correspondentes, designados de *tie points* em fotogrametria, para toda a sequência de imagens, é obtido depois de serem considerados todos os pares de imagens significativos. O conjunto

de *tie points*, juntamente com as matrizes fundamentais, são os dados de entrada requeridos para as últimas etapas da SfM (Verhoeven *et al.*, 2013).

Na fase seguinte, a triangulação (conhecida como interseção espacial direta na Fotogrametria tradicional), são calculadas as coordenadas 3D dos pontos correspondentes num sistema de coordenadas local, dando origem a uma nuvem de pontos esparsa, que representa a geometria da cena (Verhoeven *et al.*, 2013). A triangulação da imagem requer o conhecimento dos parâmetros de orientação interna da câmara e orientação externa das imagens. Estes parâmetros são obtidos após combinação de todas as orientações relativas de pares, na forma de matriz fundamental. A técnica baseia-se no Teorema Fundamental da Geometria Projetiva, que afirma que se um conjunto de pontos correspondentes em duas vistas são definidos pela matriz fundamental, então a geometria da cena 3D e as matrizes de projeção das imagens podem ser reconstruídas a partir das correspondências e quaisquer duas reconstruções vindas dessas correspondências são projetivamente equivalentes (Verhoeven *et al.*, 2013). Porém, para além de uma reconstrução projetiva, é necessária uma reconstrução métrica, ou seja, os planos ortogonais formam ângulos retos, as linhas paralelas permanecem paralelas e a reconstrução do modelo 3D é uma versão reduzida da realidade (Verhoeven *et al.*, 2013).

Na triangulação as imagens são tratadas como pares, para os quais se calculam as respetivas matrizes fundamentais. Depois de orientados os pares de imagens são combinados para formar um bloco completo de imagens e produzir a estrutura da cena. O resultado deste procedimento não é ótimo, uma vez que não são utilizadas todas as imagens sobrepostas ao mesmo tempo.

Para ultrapassar este problema, é realizado um ajustamento por feixes perspetivos. O algoritmo do ajustamento por feixes perspetivos otimiza a estrutura tridimensional e a matriz de projeção de todas as imagens simultaneamente, efetuando uma robusta minimização dos erros de medição, ou reprojeção (Verhoeven *et al.*, 2013). Por outras palavras, os feixes de raios que conectam os centros de projeção da imagem aos pontos 3D são ajustados de forma a minimizar a diferença da soma dos quadrados entre pontos de imagem observados e reprojetados. Assim, a técnica SfM consegue recuperar a geometria da cena e as matrizes de projeção da câmara através da correspondência de imagens, sem o conhecimento prévio da orientação interna, não sendo por isso necessário calibrar a câmara durante a fase de aquisição de dados, o que torna o processo muito flexível.

2.3.3 Georreferenciação

A nuvem 3D obtida após o ajustamento por feixes perspetivos é expressa num sistema de coordenadas local e equivalente ao mundo real. Para conseguir situar o modelo no mundo real é necessário utilizar dados adicionais, que definem um sistema de coordenadas de referência. Segundo Verhoeven *et al.* (2013), existem duas formas de alcançar este objetivo, que envolvem a importação de pontos fotogramétricos ou posições da câmara com elevada precisão. A primeira abordagem apresenta-se como a melhor solução, uma vez que as posições da câmara obtidas em sistemas UAV não apresentam a precisão desejada.

Para verificar a qualidade da georreferenciação, alguns PFs podem servir de pontos de controlo para comparar as coordenadas previstas com as coordenadas observadas no modelo, para esses pontos.

2.3.4 Nuvem de pontos densa

Nesta fase tem-se uma nuvem de pontos esparsa, baseada apenas em pontos de interesse. Conhecendo-se as orientações das imagens é possível criar uma nuvem tridimensional densa e texturizada. Para tal,

é executado o algoritmo MVS (*Multi-View Stereo*), que calcula uma estimativa densa da geometria observada na cena. Os algoritmos MVS recuperam a informação geométrica 3D da mesma forma que o sistema visual humano percebe a profundidade por fusão de dois pontos de vista. Quando os olhos de uma pessoa vêem um ponto numa superfície, a mente dessa pessoa combina a aparência desse ponto entre os dois pontos de vista (os dois olhos), e em seguida cruza as linhas de vista para esse ponto (triangulação) de forma a perceber a sua profundidade (Agarwal *et al.*, 2010). O MVS utiliza simultaneamente todos os pontos de vista que “vêem” um ponto para fazer uma estimativa da profundidade.

Como este algoritmo opera nos valores dos *pixels*, em vez dos pontos, permite a geração de nuvens densas de pontos a partir da nuvem esparsa inicial, o que torna possível a reconstrução dos detalhes presentes na cena (Verhoeven *et al.*, 2013).

2.3.5 Modelo 3D e ortofoto

O modelo 3D final pode ser considerado um DEM, que pode ser um DSM, ou um DTM (*Digital Terrain Model*), no caso em que apenas representa a superfície terrestre. O DSM pode ser obtido através da reconstrução de uma malha poligonal, que representa a superfície do objeto cujos vértices são os pontos da nuvem densa e neste caso é representado por uma TIN (*Triangulated Irregular Network*), ou pelo método *grid*, no qual é representado por uma grelha regular que pode ser rasterizada a partir do DSM ou da nuvem de pontos densa.

Combinando as orientações externas das imagens e interna da câmara, calculadas anteriormente, com a *grid* ou o TIN, é possível gerar ortofotomapas (imagem do bloco em que cada ponto corresponde à sua projeção ortogonal num plano horizontal, ou seja, não existem distorções características da projeção central).

3 Dados

3.1 Descrição da área de estudo

A área para a qual o método foi testado abrange aproximadamente 9,65 hectares de parte da zona norte da vila de Sousel, situada no distrito de Portalegre, no Alto Alentejo, Portugal (figura 11). Esta área é caracterizada por edificado de baixa altura, por vezes com forma complexa.

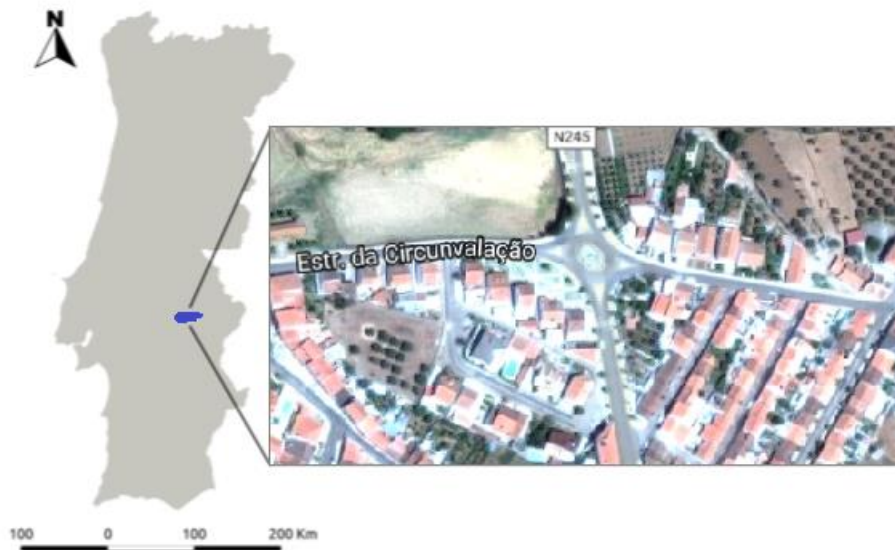


Figura 11- Enquadramento do município de Sousel no contexto nacional e área de estudo, no retângulo

3.2 Aquisição de imagens

Os dados utilizados foram obtidos num voo realizado em Janeiro de 2013, conforme descrito em Serra (2014), com a finalidade de atualizar a cartografia do município de Sousel.

O UAV utilizado foi o *swingle* CAM, da *Sensefly* (figura 12), equipado com um pequeno *chip* de GPS, que lhe confere a posição com uma exatidão que varia tipicamente entre 5 e 10 metros, baseada apenas na leitura do código de aquisição livre ou C/A (*Clear/Acquisition*); um sistema inercial IMU (*Inertial Measuring Unit*) que lhe confere os 3 ângulos de navegação (*roll, pitch e heading*), com uma exatidão que varia entre 3 a 5 graus; um rádio transmissor, com um alcance que pode chegar aos 2 Km, ou mais (dependendo da orografia do terreno) e que é usado para estabelecer a comunicação entre o *software e-motion* e o piloto automático do UAV que, por sua vez, é responsável não apenas por controlar o estado do próprio aparelho mas também por seguir um plano de voo previamente traçado, com a ajuda de sensores de pressão atmosférica e medidores da velocidade do vento (Serra, 2014). A câmara fotográfica utilizada no UAV foi a Canon IXUS 220 HS (RGB), com 12 Megapixels (4000 x 3000 pixels).



Figura 12- Aparelho UAV com o qual foram adquiridos os dados (Fonte: Serra, 2014)

O voo foi realizado com uma sobreposição longitudinal de 90% e transversal de 60%. A resolução espacial escolhida foi de 3 cm e as coordenadas dos pontos de tomada de foto encontram-se no sistema ETRS89/PT-TM06, sendo este o sistema de referência adotado para Portugal Continental.

Apenas foram utilizadas 38 das imagens obtidas durante este voo, as quais abrangem a área em destaque na figura 11. Estas imagens, juntamente com as respetivas orientações externas e interna constituem os dados primários que serão utilizados como *input* do método proposto.

A figura 13 apresenta a sobreposição de imagens obtida para o bloco.

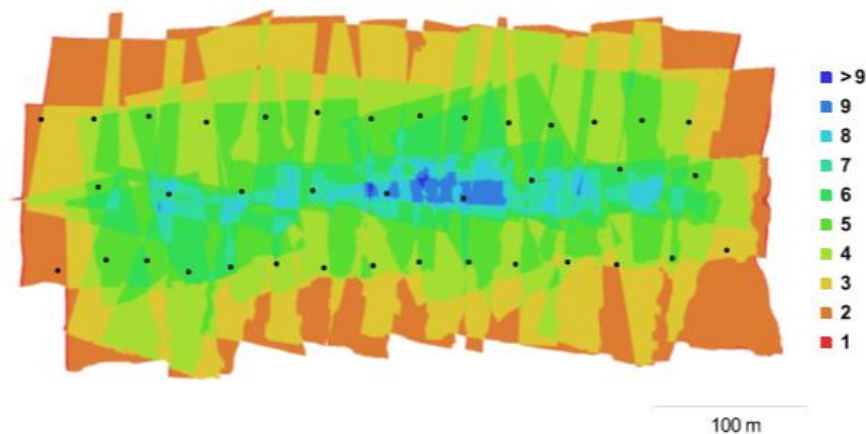


Figura 13- Localização das câmaras e sobreposição das imagens dada pelo número de fotografias em que um ponto aparece

3.3 Pontos Fotogramétricos

Para a georreferenciação do bloco, conhecem-se as coordenadas de 6 PFs (figura 14). Conforme descrito em Serra (2014), as coordenadas foram medidas com GPS em modo RTK, que proporciona uma precisão centimétrica. Podemos constatar que, para além do reduzido número de PFs, a sua distribuição pelo bloco não é a mais adequada, uma vez que se encontram todos concentrados no centro, deixando as extremidades desamparadas.



Figura 14- Distribuição de PFs no bloco

4 Operações preliminares

4.1 Criação DEM e ortofoto

O método proposto, apresentado mais à frente no capítulo 4, tem como dados de *input* não só as fotografias e as respectivas orientações externas e interna, mas também um DSM e ortofoto georreferenciados da zona coberta pelas imagens. Recorreu-se ao *software PhotoScan Agisoft* para obter estes produtos.

A tecnologia aplicada vem da área de pesquisa de *computer vision*, que desenvolve técnicas para recuperar a forma tridimensional e a aparência dos objetos em imagens, e utiliza as técnicas SfM (*Structure from Motion*) e *dense stereo-reconstruction*. Apesar de o programa ser capaz de reconstruir diversos tipos de objeto, vários testes revelam que este *software* se distingue no processamento de imagens aéreas (Verhoeven, 2011).

Após a introdução das imagens e respectivas orientações no *software* e definição do sistema de coordenadas de trabalho (neste caso o ETRS89/PT-TM06), o processo para a criação do modelo 3D e ortofoto compreende as etapas ilustradas na figura 15.

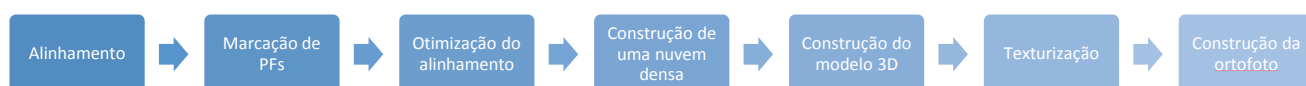


Figura 15- Workflow do *software* Agisoft PhotoScan

Durante o alinhamento, o *software* recorre à técnica SfM, através da qual consegue identificar pontos conspícuos nas imagens onde existe sobreposição, permitindo encontrar ou otimizar a posição e orientação de cada câmara e calcular ou calibrar os parâmetros de orientação interna. Desta etapa resulta uma nuvem de pontos esparsa, as posições e orientações das câmaras e os parâmetros de orientação interna.

Antes ou depois do alinhamento é importante inserir os PFs e suas coordenadas no *software*, para efetuar a georreferenciação. O mesmo PF deve ser marcado em quantas fotografias puder ser identificado (figura 16). A vantagem de realizar esta etapa depois do alinhamento é que, após introdução das coordenadas dos PFs e medição destes numa só fotografia, o *software* reconhece as restantes imagens em que ele aparece e estima a posição do mesmo nessas imagens, tornando esta tarefa mais rápida. Desta forma torna-se necessário recalcular os parâmetros de orientação interna e externa, através da otimização do alinhamento após medição de PFs.

A operação de alinhamento corresponde à determinação automática de pontos conspícuos e das orientações relativas de todas as imagens do bloco.



Figura 16- PF identificado em duas imagens diferentes

O próximo passo consiste na construção de uma nuvem de pontos densa, através do algoritmo MVS. O programa é capaz de produzir nuvens de pontos realmente muito densas, podendo mesmo ultrapassar a densidade de nuvens obtidas através do método LiDAR. A figura 17 apresenta a nuvem obtida com a qualidade máxima que o *software* permite (*ultra high*).



Figura 17- Parte Sudeste da nuvem de pontos densa com qualidade '*ultra high*'

A partir da nuvem de pontos densa, o *software* é capaz de produzir o DSM em TIN e *grid*. Ambos foram obtidos a partir da nuvem de pontos densa. O primeiro por reconstrução de uma malha poligonal e o segundo através da rasterização da nuvem, com uma resolução espacial de 3 cm/pixel. A figura 18 apresenta os dois DSM.

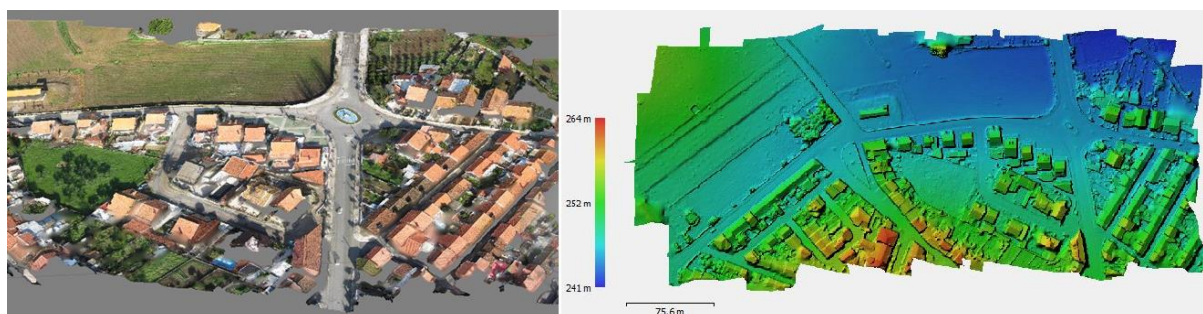


Figura 18- Parte Sudeste do DSM obtido através do método TIN (à esquerda) e através do método *grid* (à direita)

É possível atribuir textura ao modelo 3D calculando uma média ponderada do valor dos *pixels* correspondentes a determinado ponto, através da média dos vários valores possíveis, da máxima intensidade, mínima intensidade, ou pode-se ainda escolher a imagem de onde queremos recolher as texturas.

A partir do DSM em formato *grid* é então possível produzir uma ortofoto (figura 19).

O DSM (*grid*) e a ortofoto produzidos neste trabalho têm ambos uma resolução espacial de 3 cm/pixel.



Figura 19- Ortofoto

4.2 Análise da qualidade dos produtos gerados

Para avaliar a qualidade dos produtos gerados, foram definidos e analisados alguns critérios relevantes para a qualidade do produto final deste trabalho:

1. Qualidade visual da ortofoto
 - a. Nitidez e contiguidade dos contornos dos edifícios;
 - b. Geometria dos edifícios;
2. Qualidade da nuvem de pontos densa e DEM
 - a. Densidade da nuvem;
 - b. Nitidez das arestas;
 - c. Qualidade posicional da nuvem/DEM.

Sendo o objetivo deste trabalho a deteção de fachadas de edifícios e, visto que se recorre ao DEM para encontrar a posição das mesmas (tal como se explica no capítulo seguinte), a sua qualidade é um fator muito importante para o sucesso do método proposto. Neste contexto, a ortofoto tem função meramente visual, para enquadramento geográfico, pelo que a sua qualidade não tem grande influência nos resultados finais. Porém, visto que esta é obtida a partir do DEM, que é calculado com base na nuvem de pontos, podemos analisar visualmente algumas características, tais como a nitidez e contiguidade dos contornos dos edifícios e a geometria dos edifícios, que poderão ser indicadores da qualidade do DEM.

É a partir da nuvem de pontos densa que o DEM é gerado, portanto importa que esta seja o mais densa possível, para que não haja falhas nas arestas dos objetos com elevação, que são geralmente as zonas críticas neste tipo de produto.

De um modo geral a ortofoto aparenta ter boa qualidade. Porém, quando se faz *zoom* é possível verificar algumas deformações, principalmente nas zonas de menor sobreposição. A figura 20 mostra dois

exemplos em que os edifícios aparecem com uma geometria deformada e com pedaços de fachada visíveis, o que não é de esperar numa projeção ortogonal.



Figura 20- Defeitos na ortofoto

A figura 21 apresenta extratos da nuvem de pontos densa. Verifica-se que a densidade é realmente bastante elevada. Porém, existem algumas falhas, que poderão ocorrer devido à inexistência de sobreposição nessas zonas, devido a oclusões por outros telhados ou objetos nas proximidades. Ainda assim, verifica-se que, de um modo geral, as arestas dos telhados aparecem bem definidas.



Figura 21- extratos da nuvem de pontos densa

A georreferenciação foi obtida com um erro médio quadrático (EMQ) de 3,517 cm em planimetria e 0,827 cm em altimetria. O EMQ total é de 3,613 centímetros. Este valor é um indicador global da precisão da triangulação.

Não foi possível efetuar o controlo posicional ou de qualidade através de comparação dos valores obtidos com outros pontos de coordenadas conhecidas, pois o número de PFs é bastante reduzido e, portanto, optou-se por usar todos na georreferenciação do bloco. Porém, em Serra (2014), os mesmos dados foram processados num outro *software* (Pix4D) que se baseia, também ele, na técnica SfM. Nesse trabalho, o conhecimento das coordenadas de pontos de controlo, permitiu uma análise quantitativa da qualidade do posicionamento, através do cálculo do EMQ. O erro médio quadrático obtido para a altimetria (EMQ_z) foi de 68,9 cm e o erro médio quadrático obtido para a planimetria (EMQ_{xy}) foi de 38,0 cm.

A título de curiosidade, a figura 22 apresenta as elipses de erro para as posições das câmaras após a georreferenciação com PFs. Verifica-se que, de facto, a OE vinda do GPS/IMU instalado na aeronave não é suficiente para a georreferenciação das imagens, uma vez que algumas imagens chegam a ter erros na ordem do metro e meio.



Figura 22- Elipses de erro para as localizações das câmaras obtidas com GPS/IMU

5 Método proposto

O algoritmo aqui apresentado foi desenvolvido em linguagem de programação MATLAB e, como já foi referido neste documento, divide-se em duas partes: a Localização de edifícios presentes na área de estudo, consoante a morada e número de polícia, e o Reconhecimento de fachadas.

Constituem os dados de entrada as imagens, as respetivas orientações externas e interna, o DEM e a ortofoto.

O presente capítulo contém uma descrição sucinta do método.

5.1 Localização de edifícios

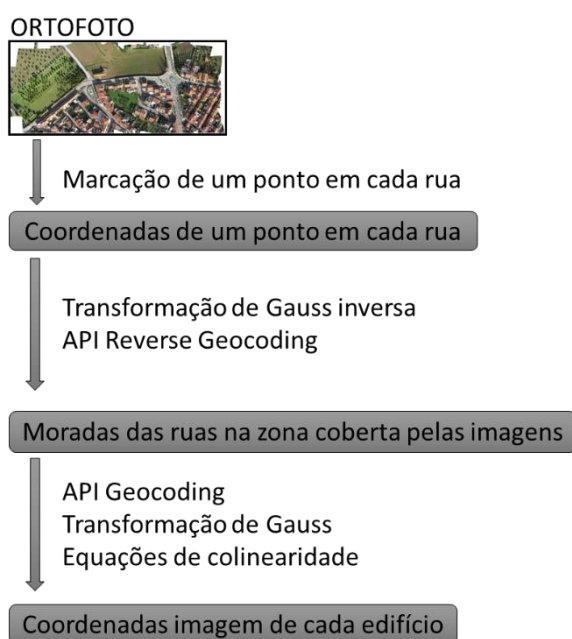


Figura 23-Fluxograma da etapa da Localização de edifícios
serviços (Google, 2016).

Como a base de dados da *Google* contém coordenadas geodésicas referentes ao *datum* WGS84 e as coordenadas dos dados de entrada são retangulares, referentes ao *datum* ETRS89, é necessário realizar uma transformação de coordenadas (transformação de Gauss) sempre que se utiliza a API. A transformação de Gauss é feita entre coordenadas do mesmo *datum* mas, como os sistemas WGS84 e ETRS89 são muito semelhantes, o efeito desta transformação será irrelevante para a precisão das coordenadas finais.

A transformação de Gauss inversa transforma as coordenadas retangulares (M, P) em coordenadas geodésicas (ϕ , λ). A dedução das fórmulas, bem como alguns conceitos teóricos acerca de projeções cartográficas podem ser consultados em Catalão (2010).

A API *Reverse Geocoding* converte uma localização numa morada. Assim, ao executar a API para cada ponto identificado pelo utilizador, é possível conhecer as moradas das ruas.

Tendo as moradas, interessa saber a localização e número de polícia dos edifícios nela presentes. O algoritmo procura a localização de edifícios iterando o número de polícia na morada até um valor considerado aceitável (tendo em conta o tamanho e características da vila de Sousel, escolheu-se o valor 100 como máximo para o valor a iterar). A cada iteração, a API *Geocoding* converte a morada na sua localização. Se o edifício se localiza na área abrangida pelas imagens, a sua localização e morada são guardadas, caso contrário, a localização é descartada.

As localizações obtidas com a API *Geocoding* encontram-se em coordenadas geodésicas. Como tal, é necessário realizar uma transformação de Gauss direta para obter as respetivas coordenadas retangulares ETRS89/PT-TM06.

É evidente que a veracidade dos dados obtidos após uso das APIs está dependente da base de dados da *Google*, que pode conter erros e lacunas. A figura 24 apresenta as localizações obtidas recorrendo aos serviços da *Google*. As APIs apenas foram executadas para as ruas na zona Sudeste do bloco. Verifica-se que muitos edifícios foram encontrados, mas existem muitos pontos que não representam nenhum edifício e alguns edifícios não foram encontrados.

Para além das APIs da *Google*, existem muitas outras que podem ser utilizadas, por exemplo, a *Bing Maps* API, a *Yahoo Maps* API, a *OS OpenSpace* API, etc. Optou-se pelo uso das APIs da *Google* devido ao número de edifícios encontrados em comparação com o número obtido com outras.



Figura 24- Resultado da aplicação da API para algumas ruas (asteriscos a vermelho)

Tendo as coordenadas planimétricas de um ponto em cada edifício, determina-se a correspondente cota para que seja possível determinar qual a imagem em que uma fachada deverá aparecer mais visível, através das equações de colinearidade. É criada uma janela de busca de 5×5 pixels, que é aplicada ao DEM, nos pontos para os quais queremos conhecer a cota. A cota do ponto no topo do edifício corresponde ao máximo encontrado no DEM, numa vizinhança com as dimensões da janela, e a cota da base corresponde ao mínimo encontrado.

Como já foi referido neste documento, as equações de colinearidade relacionam coordenadas objeto (X, Y, Z) com coordenadas imagem (x, y). Se as coordenadas x, y estiverem, respetivamente, dentro dos intervalos $[0, \text{dimX}]$ e $[0, \text{dimY}]$, em que dimX e dimY são o número de colunas e o número de linhas da imagem, em pixels, então o edifício aparece nessa imagem e resta saber se a fachada é visível nas

imagens em que o edifício aparece. O critério escolhido para saber qual a imagem ideal, da qual se vai extrair a fachada, foi a distância ao centro da imagem. Quanto maior a distância, maior será a distorção radial causada pela altura do edifício e a fachada deverá ocupar uma maior área dessa imagem. Para esta zona de estudo, este critério deverá ser suficiente, uma vez que dificilmente deverão ocorrer oclusões nas fachadas por outros objetos altos.

Para cada edifício importa distinguir a fachada Norte e Sul, ou Este e Oeste, consoante a direção de desenvolvimento da rua. Isso é determinado com base na posição do edifício na imagem (edifícios que apareçam na parte superior da imagem orientada para Norte têm a fachada Sul visível nessa imagem, edifícios na parte inferior da imagem têm a fachada Norte visível e a mesma lógica para as fachadas Este e Oeste).

5.2 Reconhecimento de fachadas

Na segunda fase é feito o reconhecimento e extração da fachada. Antes de explicar o algoritmo importa referir que nesta abordagem a fachada é vista como um plano vertical, podendo ser descrita por coordenadas 2D (x,y).

A figura 25 apresenta um fluxograma descritivo da etapa do reconhecimento de fachadas.

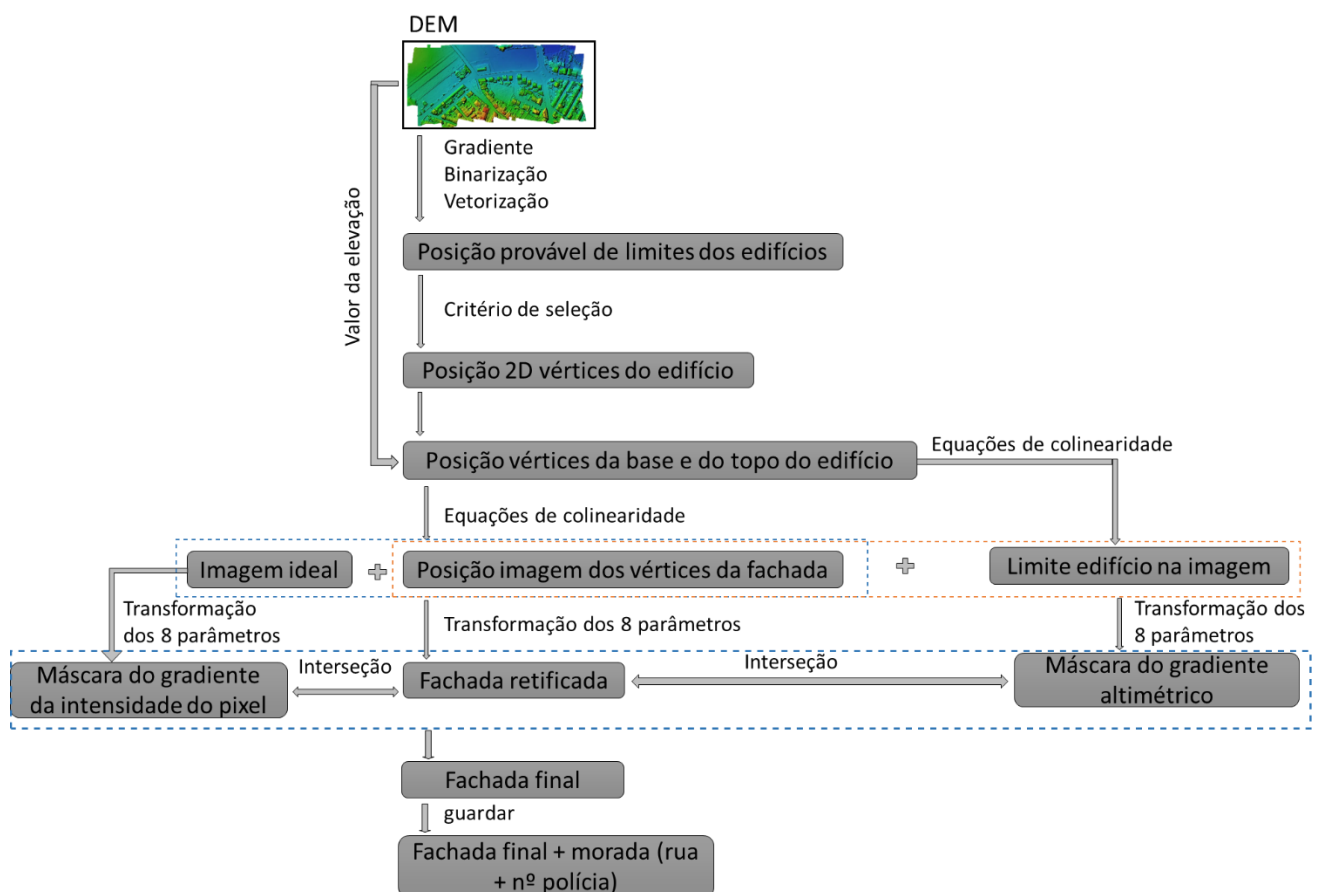


Figura 25- Fluxograma da etapa do Reconhecimento de Fachadas

O algoritmo começa por procurar as posições na imagem do DEM, onde existe probabilidade de haver uma fachada. Nos limites dos edifícios, existe uma grande diferença de altura, o que se reflete no DEM como grandes diferenças no valor do pixel que abrange o edifício, para o pixel vizinho, que não abrange o edifício. Assim sendo, calculou-se o gradiente numérico do DEM para obter estas localizações e criou-

se uma imagem binária com a representação dos *pixels* da imagem do gradiente com intensidade superior a um certo valor.

O gradiente de uma imagem dá-nos, para cada pixel, o valor da mudança direcional da intensidade do pixel. É utilizado em PDI para detetar contornos calculando o máximo e o mínimo na primeira derivada da imagem. O gradiente de uma imagem $f(x,y)$ em (x,y) é definido como:

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (8)$$

A figura 26 apresenta as posições onde existe probabilidade de existir fachadas na zona Sudeste do bloco.



Figura 26- Gradiente do DEM

Para encontrar as coordenadas dos limites de cada fachada, recorreu-se a funções criadas por Peter Kovsi, da *University of Western Australia – School of Science & software engineering*, que, a partir de uma imagem binária contendo apenas contornos, liga os pontos das arestas e cria listas de segmentos de linha, sobre as quais é possível conhecer as coordenadas dos extremos.

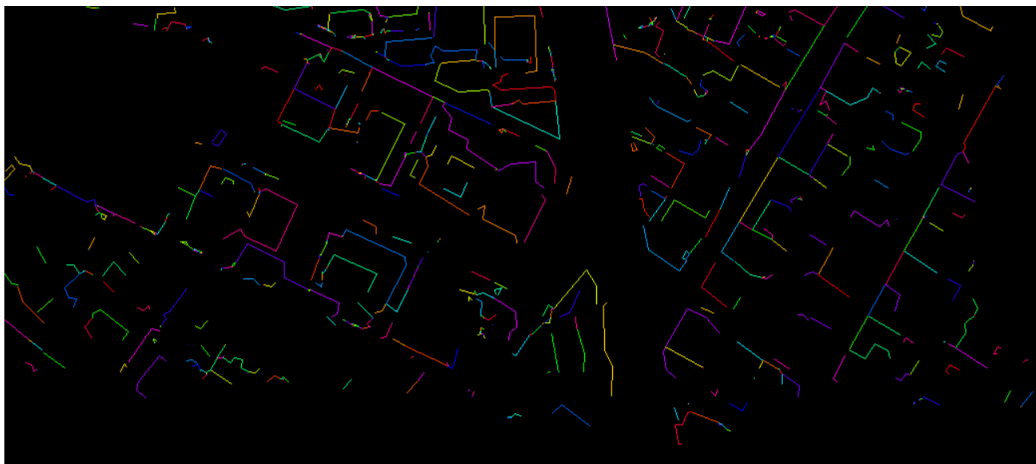


Figura 27- Posições onde existe probabilidade de existir fachadas.

A figura 27 apresenta o resultado da aplicação deste algoritmo à imagem da figura 26, a figura 28 apresenta a ortofoto com as posições onde existe probabilidade de existir fachadas e a figura 29 mostra dois zooms de partes da figura 28, para mais fácil visualização das linhas delimitadoras de edifícios.



Figura 28- Extrato da ortofoto com posições onde existe probabilidade de existir fachadas (linhas coloridas)



Figura 29- Exemplos de edifícios com respectivas posições onde existe probabilidade de existir fachada (linhas coloridas)

É possível verificar que a posição das fachadas que dão para as ruas principais aparecem bem definidas, ainda que, quando os edifícios possuem a mesma altura, o algoritmo considere as fachadas como uma só. Para resolver estes casos calculou-se o gradiente da intensidade dos *pixels* da ortofoto mas, como texturas de telhados consecutivos são, muitas vezes, bastante semelhantes, esta abordagem não teve sucesso.

Para associar as moradas obtidas na etapa da deteção de edifícios às posições 2D dos limites da fachada, o algoritmo calcula a distância mínima entre o ponto em cima do edifício obtido na etapa anterior e dois extremos de um segmento, desde que possua as condições para poder representar uma fachada (comprimento superior a 3 metros e diferença de altura numa vizinhança de (5×5) *pixels* de, pelo menos 1,8 metros - considerando possíveis erros no DEM).

De seguida, é obtida uma estimativa da coordenada Z para esses pontos 2D encontrados (extremos dos segmentos, transformando-os em coordenadas 3D dos limites na base e no topo do edifício. Através de uma pequena janela de busca de (5×5) *pixels*, o que corresponde a (15×15) centímetros no terreno, é procurado, na vizinhança de cada um dos dois pontos no DEM, o máximo e o mínimo valor do pixel

(cota). O máximo será uma aproximação à cota da posição do vértice superior da fachada e o mínimo, da posição do vértice inferior.

Tendo as coordenadas 3D dos vértices da fachada, o algoritmo calcula então as suas coordenadas imagem na imagem ideal obtida na etapa anterior, através das equações de colinearidade e as respectivas coordenadas imagem corrigidas da distorção radial e tangencial. A figura 30 apresenta um exemplo obtido para uma fachada. Os pontos a vermelho representam as posições obtidas pelo cálculo das equações de colinearidade sem os parâmetros que modelam a distorção das lentes e a verde as posições obtidas incluindo estes parâmetros. Verifica-se que as distorções radial e tangencial não afetam de forma significativa a projeção objeto-imagem.



Figura 30- Posições dos vértices de uma fachada obtidos pelo cálculo das equações de colinearidade

Seguidamente, é realizada a transformação dos 8 parâmetros para retificar a imagem da fachada, limitada pelos quatro vértices anteriores, para um plano vertical. Visto que as posições obtidas pelo cálculo das equações de colinearidade não se encontram exatamente na posição esperada, a imagem da fachada retificada irá conter *pixels* que não pertencem à fachada. O próximo passo do algoritmo é então eliminar estes *pixels*.

Para tal, criou-se uma máscara a partir do gradiente do DEM projetado na imagem (quadrilátero definido pelos pontos estimados para a posição dos vértices) e uma máscara do gradiente da intensidade dos *pixels*, que basicamente deverá detetar as arestas da fachada e todas as outras grandes alterações na cor dos *pixels* (arestas de janelas, portas e outros detalhes da fachada). No exemplo da figura 31 o limite inferior da fachada não ficou bem definido devido às semelhanças na intensidade dos *pixels* no limite fachada-chão.

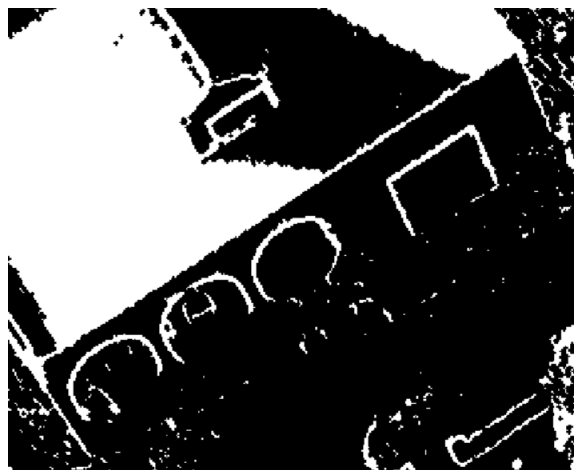


Figura 31- Gradiente da intensidade dos *pixels* para uma fachada

Projetando as duas máscaras no plano fachada e intersetando-as com a imagem da fachada retificada, obtém-se uma aproximação ao resultado final que irá conter ainda alguns elementos não pertencentes à fachada (figura 32).

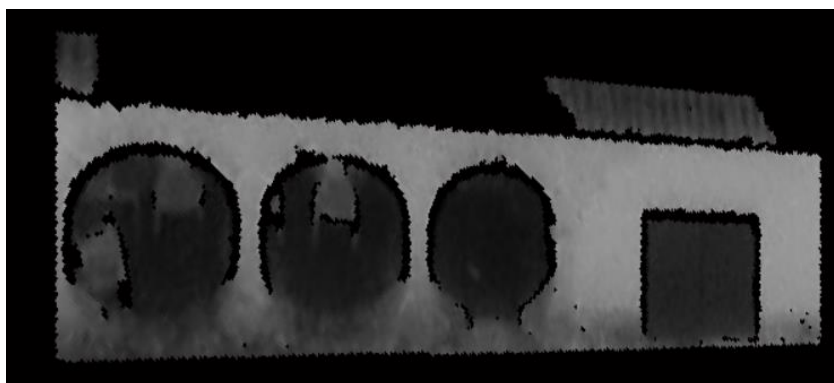


Figura 32- Fachada retificada após aplicação das máscaras

Para obter apenas a fachada são aplicados filtros morfológicos de abertura e erosão, seguidos da transformação morfológica reconstrução geodésica por dilatações sucessivas.

Os filtros morfológicos exploram as propriedades geométricas das imagens (níveis de cinza) e são representados por máscaras, denominadas elementos estruturantes, que apresentam valores 0 ou 1 na matriz que correspondem ao pixel considerado. No processo de erosão, cada pixel do objeto que possua um pixel vizinho de fundo, passa a fundo. Na dilatação, para cada pixel do objeto, todos os seus *pixels* vizinhos de fundo passam a pertencer ao objeto. A abertura é a erosão do objeto A pelo elemento estruturante B, seguido da dilatação do resultado por B. A reconstrução é o processo de realizar sucessivas dilatações sobre uma imagem Y, condicionadas por determinada geodesia X.

Após uma série de operações morfológicas, espera-se que o conjunto de *pixels* que pertence à fachada esteja isolado de outros conjuntos de *pixels* na imagem, os quais não lhe pertencem. É calculado, no espaço binário, o número de *pixels* de cada um desses conjuntos na imagem e admite-se que a fachada é representada pelo elemento com maior número de *pixels* não nulos. Para verificar se a fachada não se encontra dividida em dois ou mais elementos na imagem, compara-se o número de *pixels* do elemento de maior área, com o número de *pixels* da máscara do gradiente da elevação. Espera-se que a máscara

do gradiente altimétrico não apresente uma área muito superior à que a fachada ocupa, portanto, caso o número de *pixels* do elemento que se julga ser fachada, seja menor que dois terços do número de *pixels* não nulos da máscara do gradiente da elevação, admite-se então que a fachada é constituída pelos dois elementos de maior área da imagem. Como a fachada deve ser representada por um objeto convexo na imagem (sem *pixels* com valor nulo no interior dos limites da fachada), aplica-se uma função para tornar estes *pixels* parte da fachada, obtendo-se então a imagem da fachada. O resultado final é gravado numa diretoria, em formato .jpg, com a morada, rua e número de polícia, no nome da imagem.

A figura 33 apresenta um exemplo de uma fachada obtida pelo método.



Figura 33- Exemplo de uma fachada extraída pelo algoritmo

6 Análise da qualidade do método de reconhecimento de fachadas

O algoritmo foi testado para os 7 edifícios identificados na figura 34, que inclui parte da zona Sudeste do bloco. A escolha das fachadas a analisar teve como critério apenas a sua visibilidade nas imagens.



Figura 34- Localização dos edifícios cujas fachadas foram analisadas

Verifica-se que a precisão da posição dos vértices da fachada é muito importante para uma correta projeção no plano da fachada. Apesar de próximos da sua posição esperada, os vértices nunca se encontram na posição verdadeira. Isto poderá estar relacionado com a precisão da georreferenciação do bloco que, como se indica no capítulo 4 é de 3,5 cm em planimetria e 0,8 cm em altimetria nos PFs, o que equivale a apenas um pixel. Contudo, se tivermos em conta os erros obtidos em Serra (2014), no controlo de qualidade, o EMQ planimétrico a considerar é de 38,0 cm e o altimétrico é 68,9 cm, o que equivale a 13 *pixels* (considerando apenas o valor do EMQ planimétrico). A figura 35 apresenta um histograma da distância, em *pixels*, entre os valores calculados e os valores medidos interactivamente para as 7 fachadas.

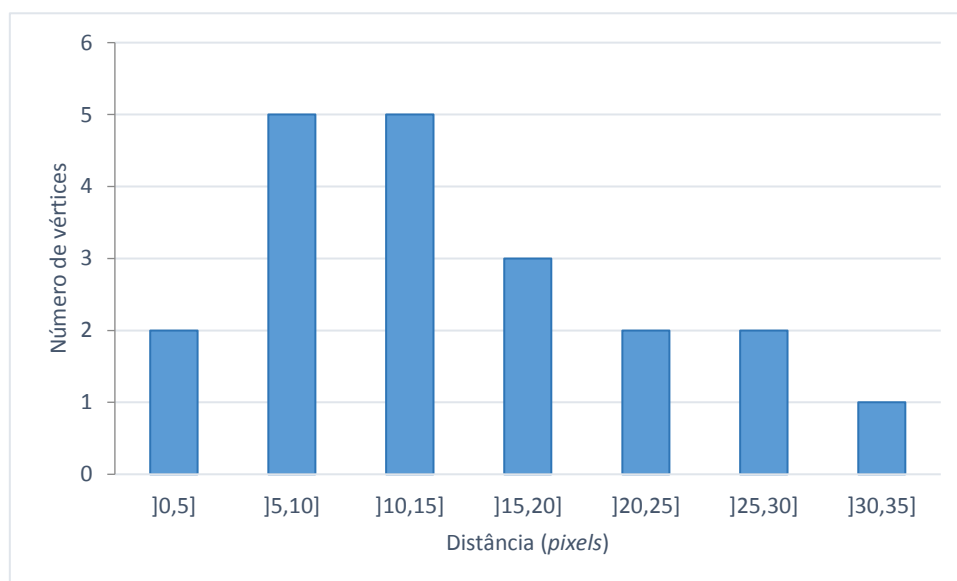


Figura 35- Histograma da distância, em pixels, entre os valores calculados e os valores medidos interactivamente, para as 7 fachadas

Para verificar a qualidade dos resultados obtidos, foi aplicada a abordagem seguida por Turker *et al.* (2016). Foi feita uma avaliação métrica quantitativa da fachada retificada. Esta métrica baseia-se na rotulagem dos *pixels* da imagem de saída com base na imagem de referência.

Consideram-se quatro categorias para atribuir ao pixel: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) e *False Negative* (FN). No caso de TP, ambos os *pixels* na imagem teste e na imagem de referência pertencem à fachada. No caso de TN, *pixels* que não pertencem à fachada são rotulados corretamente como não pertencentes à fachada. FP significa que um pixel que não pertence à fachada foi classificado como pertencente e FN significa que um pixel que pertence à fachada foi classificado como não pertencente.

Para avaliar o desempenho, foram calculadas as contagens de TP, TN, FP e FN e, em seguida, as percentagens de detecção de fachadas (pdf), a percentagem de qualidade (pq) e a percentagem de falsos negativos (pfn).

$$pdf = \frac{TP}{TP + FN} \times 100 \quad (9)$$

$$pq = \frac{TP}{TP + FP + FN} \times 100 \quad (10)$$

$$pfn = \frac{FN}{TN + FN} \times 100 \quad (11)$$

Para obter a imagem de referência, com a qual se vai comparar o resultado obtido, foi realizada a transformação dos 8 parâmetros de uma imagem com uma máscara definida pelos vértices da fachada, com coordenadas marcadas interactivamente, para o plano vertical. De seguida, retificou-se a imagem com a fachada para o mesmo plano, mas desta vez com coordenadas imagem obtidas a partir das equações de colinearidade.

Após a aplicação do algoritmo, obtém-se uma imagem com as mesmas dimensões da imagem de referência e com a fachada obtida. Efetuando a binarização dessa imagem (método de segmentação de imagens, através do qual é possível dividir uma imagem por regiões de interesse, as quais podem ser representadas a preto (valor do pixel é 0) ou branco (valor do pixel é 1)), é então possível comparar as duas imagens, e proceder às contagens de TP, TN, FP e FN.

A figura 36 apresenta a imagem binária de referência (à esquerda), e a imagem binária da fachada obtida (à direita).



Figura 36- Imagem de referência (à esquerda) e imagem obtida (à direita)

Seria de esperar que a imagem de referência fosse representada por um retângulo. Tal não acontece, pois os valores de cota são aproximações e não valores exatos.

Para obter o número de TP aplicou-se operador lógico ‘&’, que compara os *pixels* de duas imagens e retorna 1 quando o valor numa e noutra for igual, às imagens de referência e obtida e, de seguida, contaram-se os números de *pixels* com valor 1. O valor de FN foi obtido subtraindo ao número de *pixels* com valor 1 na imagem de referência, o número de TP. O número de FP corresponde ao número de *pixels* com valor 1 da imagem obtida, menos TP, e TN foi obtido subtraindo ao número total de *pixels* da imagem, o valor de TP, FN e FP.

Tabela 1- Contagens de TP, TN, FP e FN e valores obtidos para a pdf, pq e pfn

	TP	TN	FP	FN	PDF	PQ	PFN
Fachada 1	85095	270212	5310	4383	95%	90%	1,60%
Fachada 2	60536	245110	12129	3225	95%	80%	1,30%
Fachada 3	68701	267479	3797	7523	90%	86%	2,74%
Fachada 4	109352	175926	522	5200	95%	95%	2,87%
Fachada 5	55767	399598	4256	78379	42%	40%	16,40%
Fachada 6	26101	274840	6248	26811	49%	44%	8,89%
Fachada 7	115340	207401	0	111259	51%	51%	34,91%

Analisando a tabela 1, que contém os resultados obtidos no seguimento desta abordagem, é possível verificar que para alguns casos, a percentagem de deteção de fachada e de qualidade do método foram bastante satisfatórias, alcançando valores superiores a 90%. Uma percentagem de deteção de fachada muito superior à percentagem de qualidade significa que muitos *pixels* que não pertencem à fachada foram considerados como sendo fachada. Quando estes valores são muito semelhantes, o número de FP é muito reduzido. A percentagem de falsos negativos indica-nos se houve muitos *pixels* pertencentes à fachada a serem considerados como não pertencentes à mesma.

Apesar dos bons resultados para algumas fachadas, o mesmo não se verificou noutras. Analisando as características das fachadas para as quais o método fracassou, conclui-se que quanto mais detalhes a fachada tiver, e mais baixo o edifício for (provocando oclusões em zonas importantes da fachada), maior será a probabilidade de o método não ser bem sucedido.

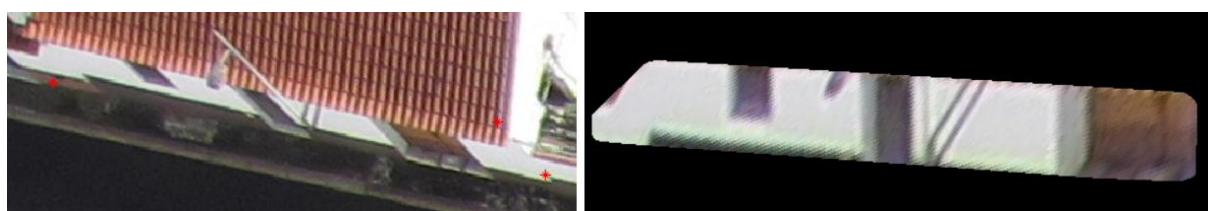


Figura 37- Exemplo de resultado para o qual o método não teve sucesso

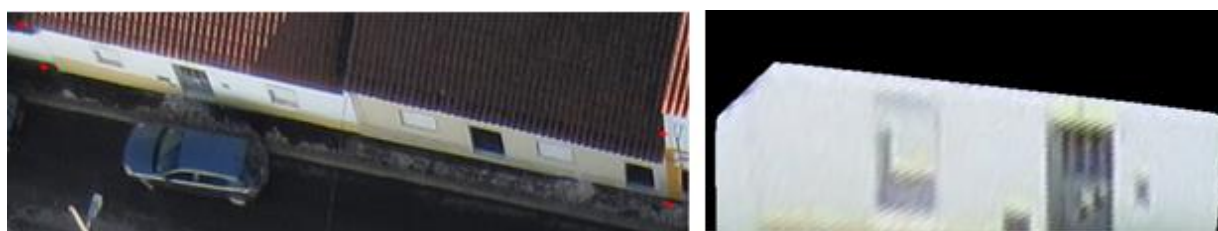


Figura 38 - Exemplo 2 de resultado para o qual o método não teve sucesso

Nos exemplos das figuras 37 e 38, o método não apresentou resultados satisfatórios. No primeiro exemplo, devido à altura do edifício e devido à perspectiva da imagem, o beiral ocultou uma boa parte da fachada, o que dificultou a interpretação do algoritmo na detecção da fachada. No segundo exemplo, duas fachadas foram reconhecidas como sendo apenas uma e o algoritmo só detetou parte de uma das fachadas. A barra amarela não foi considerada como parte da fachada.

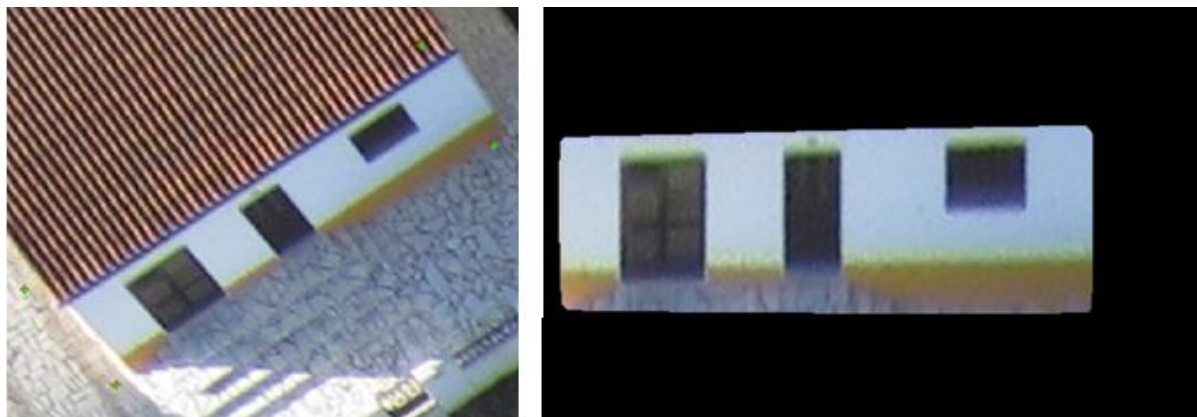


Figura 39- Exemplo de fachada obtida com sucesso



Figura 40- Exemplo de fachada obtida com sucesso

As figuras 39 e 40 apresentam fachadas obtidas com sucesso. Apesar dos detalhes na fachada, a visibilidade na imagem era boa, pelo que o algoritmo foi bem sucedido nestes casos. Verifica-se, no entanto, que uma correta posição dos vértices da fachada é bastante importante para uma boa projeção final no plano vertical. Seria de esperar que as fachadas apresentassem uma forma próxima da forma de um retângulo, o que por vezes não acontece.

A percentagem de qualidade média obtida nesta análise foi de 69% e a percentagem de deteção da fachada média foi de 74%, existindo grandes discrepâncias nestes valores, consoante as características radiométricas da fachada na imagem e também consoante a visibilidade da mesma.

A figura 41 apresenta um exemplo da aplicação das fachadas num modelo 3D, realizado no *software* CityEngine por modelação procedimental, em que as texturas são automaticamente atribuídas às fachadas dos edifícios modelados 3D, a partir de um atlas de texturas onde cada fachada está guardada num ficheiro com o nome equivalente à morada do edifício. A imagem corresponde a uma simulação.



Figura 41- Exemplo da aplicação de fachadas a um modelo 3D

7 Conclusões

As fotografias aéreas são um objeto essencial para o mapeamento urbano e contém informações sobre as fachadas dos edifícios, que podem ser extraídas. Uma vez que os modelos 3D são cada vez mais utilizados, torna-se necessário criar um método para tornar o processo de texturização o menos manual possível. Tendo em conta as características dos UAV, estes surgem como uma ferramenta de aquisição de imagens com vantagens incontestáveis (método económico, de fácil processamento e capta fachadas com boa resolução).

Neste documento apresentou-se um novo algoritmo para extração de fachadas de edifícios de imagens verticais obtidas com UAV, as quais foram adquiridas com o propósito de produzir cartografia. O método proposto é capaz de identificar os edifícios presentes na área, bem como a sua morada, estando esta tarefa, evidentemente, dependente da informação presente na base de dados da *Google*, a qual é consultada através das APIs *Reverse Geocoding* e *Geocoding*. Para além disto, é capaz de associar cada morada a uma fachada.

O método foi testado apenas para um conjunto de dados referentes à Vila de Sousel, caracterizada por edifícios de baixa altura, pelo que as fachadas não apresentam oclusões provocadas por outros edifícios. Assim, o método deverá apresentar uma taxa de sucesso inferior em zonas mais densamente urbanizadas, a não ser que se aumente a sobreposição e se utilizem fiadas cruzadas na cobertura fotográfica.

A percentagem de qualidade do método média e a percentagem de deteção de fachada média obtidas foram, respetivamente, de 69% e 74%, tendo-se obtido taxas de deteção e de qualidade superiores a 90%, para alguns casos e inferiores a 50%, noutros casos, o que põe em causa a robustez do método em algumas situações.

Os dados utilizados também foram, de certa forma, limitativos na medida em que muitas das fachadas na área de estudo apenas aparecem em uma imagem, não sendo possível explorar outras abordagens, nomeadamente a apresentada em Zebedin *et al.* (2007) que contém uma forma interessante para detetar os limites da fachada. Apesar de em algumas zonas existir uma boa sobreposição de imagens, devido à pouca área abrangida pela imagem, nalguns edifícios, as fachadas têm uma fraca visibilidade, ou nem sequer são visíveis. Também este problema poderá ser melhorado com fiadas cruzadas na cobertura fotográfica.

Também a base de dados da *Google* apresentou alguns problemas na zona testada, pois não consegue identificar muitos edifícios por não ter informação dos números de polícia. Prevê-se que este problema não seja tão grave em zonas mais urbanizadas.

Por vezes, os limites da fachada obtidos não correspondem a limites reais, provavelmente devido a imprecisões no DEM, nessa zona.

Para desenvolvimentos futuros, sugere-se a utilização de imagens oblíquas, com vista para as fachadas segundo uma melhor perspetiva. Com a câmara oblíqua a fachada deverá ser visível em mais imagens, o que permitirá expandir o leque de abordagens a enveredar. Com ou sem câmara oblíqua, sugere-se também que se tenha em atenção a importância da sobreposição de imagens. Não só para permitir a correlação da mesma fachada em imagens diferentes, mas também para permitir a construção de DEMs bem detalhados nas zonas de fachada.

8 Bibliografia

Agarwal S., Furukawa Y., Snavely N., Curless B., Seitz S. M., Szeliski R. (2010), "Reconstructing Rome. Computer", Computer, 43(6), pp.40-47.

Agisoft Photoscan software (2015), [Online]. Disponível em: <http://www.agisoft.ru>.

ANAC (2016), "Condições de operação aplicáveis aos sistemas de aeronaves pilotadas remotamente ("Drones")" - Proposta, [Online]. Disponível em: <http://www.inac.pt/vPT/Generico/LegislacaoRegulamentacao/LegislacaoConsultaPublica/HistoricoLegislacaoConsultaPublica/Paginas/HistoricoLegislacaoemConsultaPublica.aspx>. (Acedido em Novembro 2016).

Brenner C., Haala N., Fritsch D. (2001), "Towards fully automated 3D city model generation", Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images III.

Catalão J. (2010), "Projeções Cartográficas". Texto não publicado. Faculdade de Ciências da Universidade de Lisboa, Lisboa.

ESRI CityEngine (2016). [Online]. Disponível em: <http://www.esri.com/software/cityengine> (Acedido em Novembro 2016).

Fonstad M., Dietrich J., Courville B., Jensen J., Carbonneau P. (2013), "Topographic structure from motion: a new development in photogrammetric measurement", Earth Surface Processes and Landforms, 38(4), pp.421-430.

Frueh C., Sammon R., Zakhor A. (2004), "Automated Texture Mapping of 3D City Models With Oblique Aerial Imagery. 3D Data Processing, Visualization and Transmission", 3DPVT 2004. Proceedings. 2nd International Symposium on 6-9 Sept

Google (2016), Google Maps APIs. [Online]. Disponível em: <https://developers.Google.com/maps/>. (Acedido em Outubro de 2016).

Magro F. H. S. (1990), "Aerotriangulação com Métodos Alternativos na Detecção de Erros e uso de Injunções", Tese de Pós-Graduação em Ciências Geodésicas, Curitiba

Kim Z., Huertas A., Nevatia R. (2001), "Automatic description of buildings with complex rooftops from multiple images", IEEE Conference on Computer Vision and Pattern Recognition, Kauai, 2001, p. 272-279

Kosevi P. (2007), "Edge Linking and Line Segment Fitting" [Online]. Disponível em: <http://www.peterkovesi.com/matlabfns/> (Acedido em Novembro 2016).

Lowe D. (2004), "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, n.60, p.91-110.

Rau J. Y., Chu C. Y. (2012), "Photo-Realistic 3D Mapping From Aerial Oblique Imagery", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XXXVIII ISPRS Congress. Part 1.

Redweik P. (2008), Apontamentos de Fotogrametria Terrestre. Texto não publicado, Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Redweik P. (2010), Apontamentos de Fotogrametria Analítica. Texto não publicado, Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Redweik P. (2011), Apontamentos de Produção Cartográfica. Texto não publicado, Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Roberto A. J. (2013), “Extração de Informação Geográfica a partir de Fotografias Aéreas obtidas com VANTs para apoio a um SIG Municipal”, Tese de Mestrado em Sistemas de Informação Geográfica, Faculdade de Ciências do Porto, Porto.

Serra L. (2014),” O Potencial dos UAV para Atualização de Cartografia Municipal”, Tese de Mestrado em Engenharia Geográfica, Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Soares F. (2014), Apontamentos de Processamento Digital de Imagem. Texto não publicado, Faculdade de Ciências da Universidade de Lisboa, Lisboa.

Turker M., Sümer E. (2016), “Automatic near-photorealistic 3-D modelling and texture mapping for rectilinear buildings”, Geocarto International, 1-19.

Turner D., Lucieer A., Watson C. (2012), “An Automated Technique for Generating Georectified Mosaics from Ultra-High Resolution Unmanned Aerial Vehicle (UAV) Imagery, Based on Structure from Motion (SfM) Point Clouds”, Remote Sensing, 4(12), pp.1392-1410.

Verhoeven, G. (2011), “Taking computer vision aloft—archaeological three-dimensional reconstructions from aerial photographs with photostan”, Archaeological Prospection. Vol. 18. n.º 1, p. 67-73.

Verhoeven G., Christopher S., Wilfried K., Camillo R., Doneus M., Briesse C. (2013), “Undistorting the past: new techniques for orthorectification of archaeological aerial frame imagery”, Good Practice in Archaeological Diagnostics, Springer.

Westoby M., Brasington J., Glasser N., Hambrey M., Reynolds J. (2012), “Structure-from-Motion photogrammetry: A low-cost, effective tool for geoscience applications”, Geomorphology, 179, pp.300-314.

Zebedin L., Klaus A., Gruber B., Karner K. (2007), “Façade reconstruction from aerial images by multi-view plane sweeping”, PFG Vol. 2007 / 1, pp. 17 – 24.

9 Anexos

Anexo I – Exemplos de algumas fachadas obtidas com o método

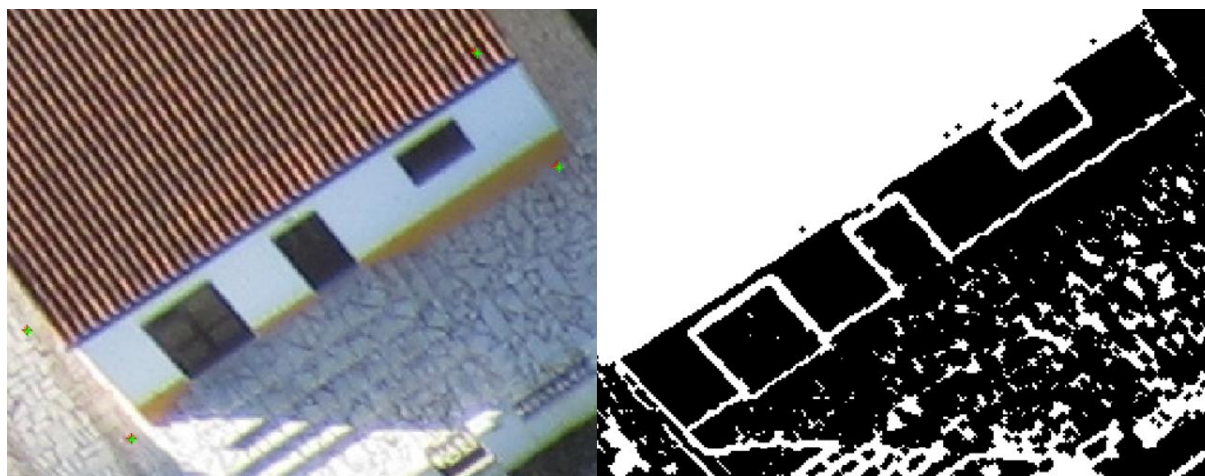


Figura 42- Estimativa da posição dos vértices da fachada 1 na imagem (à esquerda) e gradiente da intensidade dos *pixels* (à esquerda)

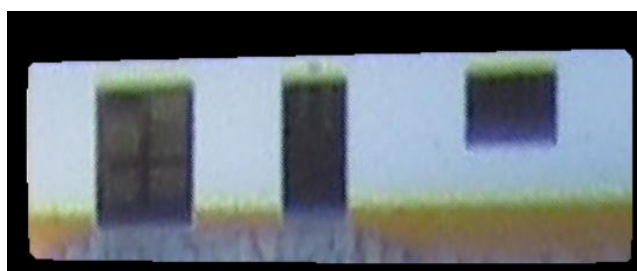


Figura 43- Exemplo 1 de fachada obtida pelo método

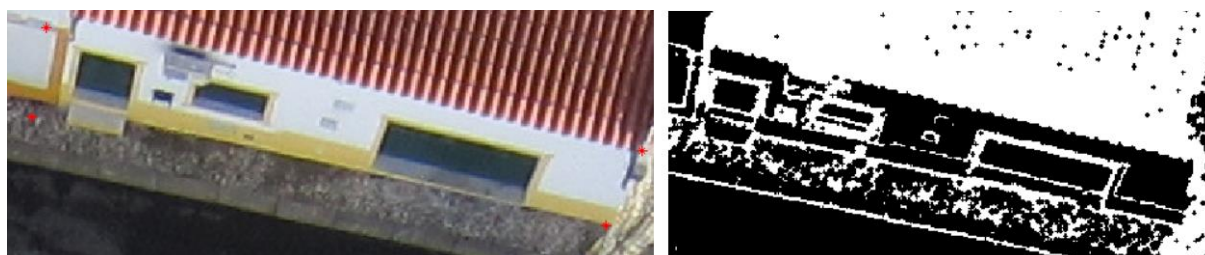


Figura 44- Estimativa da posição dos vértices da fachada 2 na imagem (à esquerda) e gradiente da intensidade dos *pixels* (à esquerda)



Figura 45- Exemplo 2 de fachada obtida pelo método



Figura 46- Estimativa da posição dos vértices da fachada 3 na imagem (à esquerda) e gradiente da intensidade dos *pixels* (à esquerda)



Figura 47- Exemplo 3 de fachada obtida pelo método



Figura 48- Estimativa da posição dos vértices da fachada 4 na imagem (à esquerda) e gradiente da intensidade dos *pixels* (à esquerda)



Figura 49- Exemplo 3 de fachada obtida pelo método